

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
11 July 2002 (11.07.2002)

(10) International Publication Number  
WO 02/053775 A2

PCT

(51) International Patent Classification<sup>7</sup>: C12Q 1/68  
24, 82347 Bernried (DE). HUSTERT, Elisabeth [DE/DE]; Geschwister Scholl Ring 2, 82110 Germering (DE).

(21) International Application Number: PCT/EP01/15290

(74) Agent: VOSSIUS & PARTNER; Siebertstrasse 4, 81675 Munich (DE).

(22) International Filing Date:  
21 December 2001 (21.12.2001)

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(30) Priority Data:  
00128627.7 28 December 2000 (28.12.2000) EP  
60/258,684 28 December 2000 (28.12.2000) US  
60/258,952 29 December 2000 (29.12.2000) US  
01100172.4 16 January 2001 (16.01.2001) EP  
60/262,859 18 January 2001 (18.01.2001) US  
01118884.4 16 August 2001 (16.08.2001) EP  
60/312,825 16 August 2001 (16.08.2001) US

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant (for all designated States except US): EPIDAUROS BIOTECHNOLOGIE AG [DE/DE]; Am Neuland 1, 82347 Bernried (DE).

Published:

— without international search report and to be republished upon receipt of that report

(72) Inventors; and  
(75) Inventors/Applicants (for US only): WOJNOWSKI, Leszek [DE/DE]; Ebenauerstrasse 9, 80637 München (DE). HABERL, Michael [DE/DE]; Waxensteinstrasse

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: IDENTIFICATION OF THE GENETIC DETERMINANTS OF THE POLYMORPHIC CYP3A5 EXPRESSION

WO 02/053775 A2

(57) Abstract: The present invention relates to a polymorphic CYP3A5 polynucleotide. Moreover, the invention relates to genes or vectors comprising the polynucleotides of the invention and to a host cell genetically engineered with the polynucleotide or gene of the invention. Further, the invention relates to methods for producing molecular variant polypeptides or fragments thereof, methods for producing cells capable of expressing a molecular variant polypeptide and to a polypeptide or fragment thereof encoded by the polynucleotide or the gene of the invention or which is obtainable by the method or from the cells produced by the method of the invention. Furthermore, the invention relates to an antibody which binds specifically the polypeptide of the invention. Moreover, the invention relates to a transgenic non-human animal. The invention also relates to a solid support comprising one or a plurality of the above mentioned polynucleotides, genes, vectors, polypeptides, antibodies or host cells. Furthermore, methods of identifying a polymorphism, identifying and obtaining a prodrug or drug or an inhibitor are also encompassed by the present invention. In addition, the invention relates to methods for producing of a pharmaceutical composition and methods of diagnosing a disease. Further, the invention relates to a method of detection of the polynucleotide of the invention. Furthermore, comprised by the present invention are a diagnostic and a pharmaceutical composition. Even more the invention relates to uses of the polynucleotides, genes, vectors, polypeptides or antibodies of the invention. Finally, the invention relates to a diagnostic kit.

**Identification of the genetic determinants of the polymorphic CYP3A5 expression**

The present invention relates to a polymorphic CYP3A5 polynucleotide. Moreover, the invention relates to genes or vectors comprising the polynucleotides of the invention and to a host cell genetically engineered with the polynucleotide or gene of the invention. Further, the invention relates to methods for producing molecular variant polypeptides or fragments thereof, methods for producing cells capable of expressing a molecular variant polypeptide and to a polypeptide or fragment thereof encoded by the polynucleotide or the gene of the invention or which is obtainable by the method or from the cells produced by the method of the invention. Furthermore, the invention relates to an antibody which binds specifically the polypeptide of the invention. Moreover, the invention relates to a transgenic non-human animal. The invention also relates to a solid support comprising one or a plurality of the above mentioned polynucleotides, genes, vectors, polypeptides, antibodies or host cells. Furthermore, methods of identifying a polymorphism, identifying and obtaining a pro-drug or drug or an inhibitor are also encompassed by the present invention. In addition, the invention relates to methods for producing of a pharmaceutical composition and to methods of diagnosing a disease. Further, the invention relates to a method of detection of the polynucleotide of the invention. Furthermore, comprised by the present invention are a diagnostic and a pharmaceutical composition. Even more, the invention relates to uses of the polynucleotides, genes, vectors, polypeptides or antibodies of the invention. Finally, the invention relates to a diagnostic kit.

The CYP3A enzymes play a particularly important role in drug metabolism. This is due to their abundant expression in the liver combined with a broad substrate spectrum. Indeed, it is estimated that CYP3A isozymes collectively comprise the largest portion of the liver CYP protein (Thummel, Annu Rev Pharmacol Toxicol 38 (1998), 389-430) and that they are involved in the metabolism of 45 % - 60 % of all currently used drugs (Li, Toxicology 104 (1995), 1-8; Evans, Science 286 (1999), 487-91). In addition to drugs, CYP3A isozymes metabolise a variety of other compounds including steroid hormones, toxins and carcinogens. For example, CYP3A isozymes metabolise aflatoxin B<sub>1</sub> (Wang, Biochemistry

37 (1998), 12536 - 45; Gillam, Arch Biochem Biophys 317 (1995), 374-84; Li, Cancer Res 57 (1997), 641-5), a mycotoxin strongly implicated in the etiology of liver cancer, which is a major cause of premature death in many areas of Africa and Asia (Henry, Science 286 (1999), 2453-4).

The hepatic expression and activity of CYP3A isozymes is inter-individually variable and this variability is the reason for harmful interactions frequently encountered in development and application of drugs that are CYP3A substrates. It has also been postulated that variable CYP3A expression could affect an individual's predisposition to cancers caused by environmental carcinogens which are metabolised by CYP3A. The elucidation of factors controlling an individual's CYP3A activity could permit personalised dose adjustments in therapies with its substrates and also lead to the identification of sub-populations at increased risk for several common cancers. However, despite considerable efforts, our understanding of factors governing CYP3A activity and expression is limited. There are several reasons for this: An average human liver may express products of up to four CYP3A genes (Gellner, Pharmacogenetics 11 (2001), 111 - 121), but their respective contributions to the hepatic CYP3A pool are still a matter of debate. The differentiation between the individual CYP3A proteins by enzymatic methods has proven difficult due to overlapping substrate specificities and due to the considerable effect of reconstitution conditions on their catalytic activities. RNA and protein analysis indicate that CYP3A4 forms the bulk of the hepatic CYP3A protein and its expression is highly variable (Thummel, Annu Rev Pharmacol Toxicol 38 (1998), 389-430). Less well understood are the contributions of the other CYP3A genes. CYP3A5 is widely considered the second most important CYP3A protein in the liver, but the available data are conflicting, since its expression has been reported to be present in 10 % to 97 % of human livers (Aoyama, J Biol Chem 264 (1989), 10388-95; Wrighton, Mol Pharmacol 38 (1990), 207-13; Schuetz, Pharmacogenetics 4 (1994), 11-20; Jounaidi, Biochem Biophys Res Commun 221 (1996), 466-70; Boobis, Br J Clin Pharmacol 42 (1996), 81-9). The possible reasons for these discrepancies include small sample sizes, interethnic differences and poor specificity of probes used to measure CYP3A5 expression. The third CYP3A, CYP3A7, was originally described in the human fetal liver where it accounts for about 50 % of the total CYP protein (Wrighton, Biochem Pharmacol 37 (1988), 3053-5). More recent studies indicate constitutive or induced expression of CYP3A7 in adult human livers, but its quantification has been hampered by the lack of specific antibodies. Similarly, no protein expression

data are available for the recently identified fourth member of the family, CYP3A43 (Gellner, Pharmacogenetics 11 (2001), 111 - 121).

Clinical studies indicate that a major portion of the inter-individual CYP3A variability is caused by genetic factors (Ozdemir, Pharmacogenetics 10 (2000), 373-88), but the identities of the latter remain unknown. In respect of CYP3A5, a protein variant (Thr398Asn) has been found in 2 out of 5 individuals deficient in CYP3A5 expression (Jounaidi, Biochem Biophys Res Commun 221 (1996), 466-70), but its significance has not been verified on a larger number of liver samples and in functional studies. In addition, a haplotype consisting of two linked polymorphisms has been described in the 5' flanking region of the CYP3A5 gene which is associated with increased expression and activity of the gene (Paulussen, Pharmacogenetics 10 (2000), 415-24). However, only a small sample set (n=29) was analysed for the genotype and the phenotype. Moreover, the single nucleotide polymorphisms (SNPs) which have been disclosed in said document are not suitable for a reliable prediction of CYP3A5 dysfunction and/or dysregulation and the problems caused thereby. This document does not suggest the existence of further haplotypes.

Thus, improved means and methods for diagnosing and treating a variety of diseases and disorders based on dysfunctions or dysregulations of drug metabolism were not available yet but are nevertheless highly desirable. Thus, the technical problem underlying the present invention is to comply with the above specified needs.

The solution to this technical problem is achieved by providing the embodiments characterized in the claims.

Accordingly, the present invention relates to a polynucleotide comprising a polynucleotide selected from the group consisting of:

- (a) a polynucleotide having the nucleic acid sequence of SEQ ID NO: 54, 56, 58, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 106, 108, 110, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 128, 129, 130, 131, 133, 134, 135, 136, 137, 138, 139, 140, 142, 143, 149, 151, 153, 155, 157, 159, 161, 163, 165, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 193, 195, 197, 199, 201, 207, 208,

209, 210, 211, 212, 213, 214, 216, 218, 219, 220, 221, 222, 223, 224, 225,  
226, 227, 228, 231, 232, 233, 235, or 236;

- (b) a polynucleotide encoding a polypeptide having the amino acid sequence of SEQ ID NO: 127, 132, 141, 215, 229, or 234;
- (c) a polynucleotide capable of hybridizing to a CYP3A5 gene, wherein said polynucleotide is having a nucleotide exchange, a nucleotide deletion of at least one nucleotide, or at least one additional nucleotide at a position corresponding to position -20643, -20555, -20359, -20367, -20329, -20323, -20310, -6200, -6177, -4336, -3990, -3868, -3844, -3557, -1617, -795, -86, -74, 136, 174 to 176, 230, 3705, 3709/3710, 5215, 5235, 5516, 7182, 7207, 7303, 7424/7427, 12907, 13028, 13077, 13173, 13226, 13376, 14720, 14836, 14903, 15788, 16079, 16931/16932, 16993, 17163, 19069, 19165, 19208, 27050, 27131/27132, 27526, 31499, 31551 or 31611 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614);
- (d) a polynucleotide capable of hybridizing to a CYP3A5 gene, wherein said polynucleotide is having an A at a position corresponding to position -20555, -20329, -20323, -4336, -3868, -3844, -795, -86, 230, 5235, 5516, 7182, 7303, 12907, 13028, 13376, 19069 or 19165 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614), a T at a position corresponding to position -20367, -6200, -74, 3705, 5215, 7207, 14836, 17163, 19208 or 27526 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614), a G at a position corresponding to position -6177, -3990, 13077, 14720, 14903, 16993 or 27050 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614), a C at a position corresponding to position -20643, -20310, -3557, -1617, 136, 13173, 13226, 15788, 16079, 31499, 31551 or

31611 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614), nucleotide deletions at positions corresponding to positions 174 to 176 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613), an additional nucleotide at a position corresponding to position 3709/3710 or 27131/27132 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614), three additional nucleotides at a position corresponding to position 16931/16932 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614), or a deletion of two nucleotides and nine additional nucleotides inserted at a position corresponding to position 7424 to 7427 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613);

- (e) a polynucleotide encoding a CYP3A5 polypeptide or fragment thereof, wherein said polypeptide comprises an amino acid substitution at a position corresponding to position 30, 100, 130, 149 or 488 of the CYP3A5 polypeptide (Accession No: NP\_000768.1), or at least one amino acid exchange or a stop codon at a position corresponding to position 30 to 34 or 346 to 348 of the CYP3A5 polypeptide (Accession No: NP\_000768.1); and
- (f) a polynucleotide encoding a CYP3A5 polypeptide or fragment thereof, wherein said polypeptide comprises amino acid substitutions of HGLFK to YGTF. (with the period meaning termination) at a position corresponding to position 30 to 34 of the CYP3A5 polypeptide (Accession No: NP\_000768.1, an amino acid substitution of S to Y at a position corresponding to position 100 of the CYP3A5 polypeptide (Accession No: NP\_000768.1), an amino acid substitution of R to Q at a position corresponding to position 130 of the CYP3A5 polypeptide (Accession No: NP\_000768.1), an amino acid substitution of I to T at a position corresponding to position 149 of the CYP3A5 polypeptide (Accession No: NP\_000768.1), an amino acid

substitutions of TYD to YL. (with the period meaning termination) at position corresponding to position 346 to 348 of the CYP3A5 polypeptide (Accession No: NP\_000768.1), or an amino acid substitution of I to T at a position corresponding to position 488 of the CYP3A5 polypeptide (Accession No: NP\_000768.1).

In the context of the present invention the term "polynucleotides" or the term "polypeptides" refers to different variants of a polynucleotide or polypeptide. Said variants comprise a reference or wild type sequence of the polynucleotides or polypeptides of the invention as well as variants which differ therefrom in structure or composition. Reference or wild type sequences for the polynucleotides are Accession No: AF280107.1 and AC005020.2. Reference or wild type sequence for the polypeptides of the invention is Accession No: NP\_000768.1. The differences in structure or composition usually occur by way of nucleotide or amino acid substitution(s), addition(s) and/or deletion(s). Preferably, said nucleotide substitution(s), addition(s) or deletion(s) result(s) in one or more changes of the corresponding amino acid(s) of the polypeptides of the invention. The variant polynucleotides and polypeptides also comprise fragments of said polynucleotides or polypeptides of the invention. The polynucleotides and polypeptides as well as the aforementioned fragments thereof of the present invention are characterized as being associated with a CYP3A5 dysfunction or dysregulation. Preferably, said dysfunctions or dysregulations referred to in the present invention cause a disease or disorder or a prevalence for said disease or disorder. Preferably, as will be discussed below in detail, said disease is cancer or diseases including cardiovascular diseases, diabetes and AIDS or any other disease caused by a dysfunction or dysregulation due to a polynucleotide or polypeptides of the invention.

The polynucleotides of the invention include polynucleotides that have at least 70%, preferably at least 75%, at least 80%, at least 85%, at least 90% or at least 95% sequence identity to a CYP3A5 gene, wherein said polynucleotide is having a nucleotide exchange, a nucleotide deletion of at least one nucleotide, or at least one additional nucleotide at a position corresponding to position -20643, -20555, -20359, -20367, -20329, -20323, -20310, -6200, -6177, -4336, -3990, -3868, -3844, -3557, -1617, -795, -86, -74, 136, 174 to 176, 230, 3705, 3709/3710, 5215, 5235, 5516, 7182, 7207, 7303, 7424/7427, 12907, 13028, 13077, 13173, 13226, 13376, 14720, 14836, 14903, 15788, 16079, 16931/16932, 16993, 17163, 19069, 19165, 19208, 27050, 27131/27132, 27526, 31499, 31551 or 31611

of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614).

The term "hybridizing" as used herein refers to polynucleotides which are capable of hybridizing to the polynucleotides of the invention or parts thereof which are associated with a CYP3A5 dysfunction or dysregulation. Thus, said hybridizing polynucleotides are also associated with said dysfunctions and dysregulations. Therefore, said polynucleotides may be useful as probes in Northern or Southern Blot analysis of RNA or DNA preparations, respectively, or can be used as oligonucleotide primers in PCR analysis dependent on their respective size. Also comprised by the invention are hybridizing polynucleotides which are useful for analysing DNA-Protein interactions via, e.g., electrophoretic mobility shift analysis (EMSA). Preferably, said hybridizing polynucleotides comprise at least 10, more preferably at least 15 nucleotides in length while a hybridizing polynucleotide of the present invention to be used as a probe preferably comprises at least 100, more preferably at least 200, or most preferably at least 500 nucleotides in length.

It is well known in the art how to perform hybridization experiments with nucleic acid molecules, i.e. the person skilled in the art knows what hybridization conditions s/he has to use in accordance with the present invention. Such hybridization conditions are referred to in standard text books such as Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory (1989) N.Y. Preferred in accordance with the present inventions are polynucleotides which are capable of hybridizing to the polynucleotides of the invention or parts thereof which are associated with a CYP3A5 dysfunction or dysregulation under stringent hybridization conditions, i.e. which do not cross hybridize to unrelated polynucleotides such as polynucleotides encoding a polypeptide different from the CYP3A5 polypeptides of the invention.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, or organic solvents, in addition to the base composition, length of the complementary strands and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. Stringent temperature conditions will generally include temperatures in excess of 30°C, typically 37°C, and preferably in excess of 45°C. Stringent salt conditions will ordinarily be less than 1000 mM, typically less than 500 mM and preferably less than 200 mM. However, the combination of parameters is much more important than the measure of any single parameter. See, e.g., Wetmur and Davidson, 1968. Probe sequences may also hybridize

specifically to duplex DNA under certain conditions to form triplex or higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well known in the art.

The term "percent sequence identity" or "identical" in the context of nucleic acid sequences refers to the residues in the two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least nine nucleotides, usually at least 20 nucleotides, more usually at least 24 nucleotides, typically at least 28 nucleotides, more typically at least 32 nucleotides, and preferably at least 36 nucleotides or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using Fasta, a program in GCG Version 6.6, Fasta provides alignments and percent sequence identity of the regions of the best overlap between the query and the search sequence (Pearson, 1980, herein incorporated by reference). For instance, percent sequence identity between nucleic acid sequences can be determined using Fasta with its default parameters (a word size of 6 and the NOPAMfactor for the scoring matrix) as provided in GCG Version 6.1, herein incorporated by reference.

The term "corresponding" as used herein means that a position is not only determined by the number of the preceding nucleotides and amino acids, respectively. The position of a given nucleotide or amino acid in accordance with the present invention which may be deleted, substituted or comprise one or more additional nucleotide(s) may vary due to deletions or additional nucleotides or amino acids elsewhere in the gene or the polypeptide. Thus, under a "corresponding position" in accordance with the present invention it is to be understood that nucleotides or amino acids may differ in the indicated number but may still have similar neighboring nucleotides or amino acids. Said nucleotides or amino acids which may be exchanged, deleted or comprise additional nucleotides or amino acids are also comprised by the term "corresponding position". Said nucleotides or amino acids may for instance together with their neighbors form sequences which may be involved in the regulation of gene expression, stability of the corresponding RNA or RNA editing, as well as encode functional domains or motifs of the protein of the invention.

By, e.g., "position 3709/3710" it is meant that said polynucleotide comprises one or more additional nucleotide(s) which are inserted between positions 3709 and position 3710 of

the corresponding wild type version of said polynucleotide. The same applies mutatis mutandis to all other position numbers referred to in the above embodiment which are drafted in the same format, i.e. two consecutive position numbers separated by a slash (/). By, e.g., "position 7424 to 7427" is meant that said polynucleotide comprises one or more deleted nucleotides which are deleted between positions 7424 and position 7427 of the corresponding wild type version of said polynucleotide and/or one or more additional nucleotide(s) which are inserted between positions 7424 and position 7427 of the corresponding wild type version of said polynucleotide. The same applies mutatis mutandis to all other position numbers referred to in the above embodiment which are drafted in the same format.

The numbering of the polymorphisms refers to the aligned and joined genomic sequences AF280107.1 and AC005020.2, wherein the T at position 174832 (which has been numbered +8613) of the sequence AF280107.1 refers to position 27340 of the sequence AC005020.2. The nucleotide A at position 27341 of the sequence AC005020.2 has been numbered +8614. Numbering of polymorphisms to a position corresponding to a position up to +8613 refers to the genomic sequence AF280107.1, numbering of polymorphisms to a position corresponding to position +8614 and greater refer to the genomic sequence AC005020.2.

In accordance with the present invention, the mode and population distribution of genetic variations in the CYP3A5 gene has been analyzed by sequence analysis of relevant regions of the human said gene from many different individuals. It is a well known fact that genomic DNA of individuals, which harbor the individual genetic makeup of all genes, including the CYP3A5 gene, can easily be purified from individual blood samples. These individual DNA samples are then used for the analysis of the sequence composition of the alleles of the CYP3A5 gene that are present in the individual which provided the blood sample. The sequence analysis was carried out by PCR amplification of relevant regions of said genes, subsequent purification of the PCR products, followed by automated DNA sequencing with established methods (e.g. ABI dyeterminator cycle sequencing).

One important parameter that had to be considered in the attempt to determine the individual genotypes and identify novel variants of the CYP3A5 gene by direct DNA-sequencing of PCR-products from human blood genomic DNA is the fact that each human harbors (usually, with very few abnormal exceptions) two gene copies of each autosomal gene (diploidy). Because of that, great care had to be taken in the evaluation of the sequences to be able to identify unambiguously not only homozygous sequence variations

but also heterozygous variations. The details of the different steps in the identification and characterization of novel polymorphisms in the CYP3A5 gene (homozygous and heterozygous) are described in the Examples below.

Over the past 20 years, genetic heterogeneity has been increasingly recognized as a significant source of variation in drug response. Many scientific communications (Meyer, Ann. Rev. Pharmacol. Toxicol. 37 (1997), 269-296 and West, J. Clin. Pharmacol. 37 (1997), 635-648) have clearly shown that some drugs work better or may even be highly toxic in some patients than in others and that these variations in patient's responses to drugs can be related to molecular basis. This "pharmacogenomic" concept spots correlations between responses to drugs and genetic profiles of patient's (Marshall, Nature Biotechnology, 15 (1997), 954-957; Marshall, Nature Biotechnology, 15 (1997), 1249-1252). In this context of population variability with regard to drug therapy, pharmacogenomics has been proposed as a tool useful in the identification and selection of patients which can respond to a particular drug without side effects. This identification/selection can be based upon molecular diagnosis of genetic polymorphisms by genotyping DNA from leukocytes in the blood of patient, for example, and characterization of disease (Bertz, Clin. Pharmacokinet. 32 (1997), 210-256; Engel, J. Chromatogra. B. Biomed. Appl. 678 (1996), 93-103). For the founders of health care, such as health maintenance organizations in the US and government public health services in many European countries, this pharmacogenomics approach can represent a way of both improving health care and reducing overheads because there is a large cost to unnecessary drugs, ineffective drugs and drugs with side effects.

The mutations in the variant genes of the invention sometime result in amino acid deletion(s), insertion(s) and in particular in substitution(s) either alone or in combination. It is of course also possible to genetically engineer such mutations in wild type genes or other mutant forms. Methods for introducing such modifications in the DNA sequence of said genes are well known to the person skilled in the art; see, e.g., Sambrook, Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory (1989) N.Y.

For the investigation of the nature of the alterations in the amino acid sequence of the polypeptides of the invention software programs may be used such as RASMOL that are obtainable from the Internet. Furthermore, folding simulations and computer redesign of structural motifs can be performed using other appropriate computer programs (Olszewski,

Proteins 25 (1996), 286-299; Hoffman, Comput. Appl. Biosci. 11 (1995), 675-679). Computers can be used for the conformational and energetic analysis of detailed protein models (Monge, J. Mol. Biol. 247 (1995), 995-1012; Renouf, Adv. Exp. Med. Biol. 376 (1995), 37-45). These analysis can be used for the identification of the influence of a particular mutation on binding and/or processing of drugs.

Usually, said amino acid deletion, addition or substitution in the amino acid sequence of the protein encoded by the polynucleotide of the invention is due to one or more nucleotide substitution, insertion or deletion, or any combinations thereof. Preferably said nucleotide substitution, insertion or deletion may result in amino acid substitutions of HGLFK to YGTF. (with the period meaning termination) at a position corresponding to position 30 to 34 of the CYP3A5 polypeptide (Accession No: NP\_000768.1), in an amino acid substitution of S to Y at a position corresponding to position 100 of the CYP3A5 polypeptide (Accession No: NP\_000768.1), in amino acid substitutions of TYD to YL. (with the period meaning termination) at a position corresponding to position 346 to 348 of the CYP3A5 polypeptide (Accession No: NP\_000768.1), or in an amino acid substitution of T to N at a position corresponding to position 398 of the CYP3A5 polypeptide (Accession No: NP\_000768.1).

The mutations in the CYP3A5 gene detected in accordance with the present invention are listed in Table 2A-E. The methods of the mutation analysis followed standard protocols and are described in detail in the Examples. In general such methods are to be used in accordance with the present invention for evaluating the phenotypic spectrum as well as the overlapping clinical characteristics of diseases or conditions related to dysfunctions and diseases related to the drug metabolism. Advantageously, the characterization of said mutants may form the basis of the development of improved drugs, such as drugs which are used e.g. in cancer therapy and diseases including cardiovascular diseases, diabetes and AIDS. Said methods encompass for example haplotype analysis, single-strand conformation polymorphism analysis (SSCA), PCR and direct sequencing, or TaqMan® analysis. On the basis of thorough clinical characterization of many patients the phenotypes can then be correlated to these mutations as well as to mutations that had been described earlier, for example in Jounaidi, Biochem Biophys Res Commun, 221, pp. 466-470, 1996.

Also comprised by the polynucleotides referred to in the present invention are polynucleotides which comprise at least two, preferably at least three, of the polynucleotides specified hereinabove, i.e. polynucleotides having a nucleotide sequence which contains at least two, preferably three of the mutations comprised by the above polynucleotides or listed in the tables below. Thus, the haplotype determined in accordance with the present invention can be characterized by at least two, preferably three of said mutations in the CYP3A5 locus. Further, the polynucleotide of the invention may further comprise at least one nucleotide deletion, addition and/or substitution other than those specified hereinabove, for example those described in the prior art; e.g., in Jounaidi, Biochem Biophys Res Commun, 221, pp. 466-470, 1996, in Paulussen, Pharmacogenetics 10, pp. 415-424, 2000, in Kuehl, 2001, Nature Genetics 27: 383-391, or in Chou, 2001, Drug Metab Dispos 29: 1205-1209.

This allows the study of synergistic effects of said mutations in the CYP3A5 gene and/or a polypeptide encoded by said polynucleotide on the pharmacological profile of drugs in patients who bear such mutant forms of the gene or similar mutant forms that can be mimicked by the above described proteins. It is expected that the analysis of said synergistic effects provides deeper insights into the onset of dysfunctions or diseases related to drug metabolism as described supra. From said deeper insight the development of diagnostic and pharmaceutical compositions related to dysfunctions or diseases related to drug metabolism will greatly benefit.

Moreover, it has been surprisingly found that the so called positive predictive power for CYP3A5 dysfunctions or dysregulations can be significantly increased based on the polynucleotides of the present invention and thus allows a reliable prediction in contrast to positive predictive power based on the prior art. The increased CYP3A5 protein expression in all except one liver samples (17/18) identified in accordance with the present invention and described in detail in the examples below co-segregates with a haplotype which consists of at least three variants (ch-v-021, ch-v-026, ch-v-015) with distinct locations within or upstream of the gene locus. Genotyping these three variants has in no case led to the generation of false-positive predictions resulting in an estimated positive predictive power for the 3-variant genotype of about 99.95 %. This is in striking contrast to the positive predictive power determined for the haplotype described by Paulussen, Pharmacogenetics 10, pp. 415-424, 2000 which is about 65 %. Moreover, based on the polynucleotides of the invention and as described in the examples below, it has been found that the SNPs described by Paulussen, Pharmacogenetics 10, pp. 415-424, 2000

are located in contrary to what is reported in said document approximately 20 kb upstream of the transcriptional start site of the CYP3A5 gene in a sequence 5' to a CYP3A5 pseudogene locus.

Therefore, the haplotypes characterized on the basis of the polynucleotides of the present invention fulfil the criteria expected from a reliable marker of CYP3A5 expression. As is evident to the person skilled in the art, the genetic knowledge deduced from the present invention can now be used to exactly and reliably characterize the genotype of a patient. Advantageously, diseases or a prevalence for a disease which are associated with CYP3A5 dysfunction or dysregulation, such as cancer, diseases including cardiovascular diseases, diabetes and AIDS, can be predicted and preventive or therapeutical measures can be applied accordingly. Moreover in accordance with the foregoing, in cases where a given drug takes an unusual effect, a suitable individual therapy can be designed based on the knowledge of the individual genetic makeup of a subject with respect to the polynucleotides of the invention and improved therapeutics can be developed as will be further discussed below.

Finally, the polynucleotides and polypeptides referred to in accordance with the present invention are also useful as forensic markers, which improve the identification of subjects which have been murdered or killed by, for example, a crime of violence or any other violence and can not be identified by the well known conventional forensic methods. The application of forensic methods based on the detection of the polymorphisms comprised by the polynucleotides of this invention in the genome of a subject are particularly well suited in cases where a (dead) body is disfigured in a severe manner such that identification by other body characteristics such as the features of the face is not possible. This is the case, for example, for corpse found in water which are usually entirely disfigured. Advantageously methods which are based on the provision of the polynucleotides of the invention merely require a minimal amount of tissues or cells in order to be carried out. Said tissues or cells may be blood droplets, hair roots, epidermal scales, saliva droplets, sperms etc. Since only such a minimal amount of tissues or cells are required for the identification of a subject, the polymorphisms comprised by the polynucleotides of this invention can be also used as forensic markers in order to prove someone guilt of a crime, such as a violation or a ravishment. Moreover, the polymorphisms comprised by the polynucleotides of this invention can be used to proof paternity. In accordance with the forensic methods referred to herein the presence or absence of the polynucleotides of the invention is determined and compared with a reference sample which is unambiguously derived from the subject to be identified. The

forensic methods which require detection of the presence or absence of the polynucleotides of the invention in a sample of a subject the polymorphisms comprised by the polynucleotides of this invention can be for example PCR-based techniques which are particularly well suited in cases where only a minimal amount of tissues or cells are available as forensic samples. On the other hand, where enough tissue or cells are available, hybridization based techniques may be performed in order to detect the presence or absence of a polynucleotide of this invention. These techniques are well known by the person skilled in the art and can be adopted to the individual purposes referred to herein without further ado. In conclusion, thanks to the present invention forensic means which allow improved and reliable predictions as regards the aforementioned aspects are now available...

In line with the foregoing, preferably, the polynucleotide of the present invention is associated with cancer or diseases including cardiovascular diseases, diabetes and AIDS.

The term "cancer" used herein is very well known and characterized in the art. Several variants of cancer exist and are comprised by said term as meant in accordance with the invention. For a detailed list of symptoms which are indicative for cancer it is referred to text book knowledge, e.g. Pschyrembel.

In a further embodiment the present invention relates to a polynucleotide which is DNA or RNA.

The polynucleotide of the invention may be, e.g., DNA, cDNA, genomic DNA, RNA or synthetically produced DNA or RNA or a recombinantly produced chimeric nucleic acid molecule comprising any of those polynucleotides either alone or in combination. Preferably said polynucleotide is part of a vector, particularly plasmids, cosmids, viruses and bacteriophages used conventionally in genetic engineering that comprise a polynucleotide of the invention. Such vectors may comprise further genes such as marker genes which allow for the selection of said vector in a suitable host cell and under suitable conditions.

The invention furthermore relates to a gene comprising the polynucleotide of the invention.

It is well known in the art that genes comprise structural elements which encode an amino acid sequence as well as regulatory elements which are involved in the regulation of the expression of said genes. Structural elements are represented by exons which may either encode an amino acid sequence or which may encode for RNA which is not encoding an amino acid sequence but is nevertheless involved in RNA function, e.g. by regulating the stability of the RNA or the nuclear export of the RNA.

Regulatory elements of a gene may comprise promoter elements or enhancer elements both of which could be involved in transcriptional control of gene expression. It is very well known in the art that a promoter is to be found upstream of the structural elements of a gene. Regulatory elements such as enhancer elements, however, may be found distributed over the entire locus of the gene. Said elements could be reside, e.g., in introns, regions of genomic DNA which separate the exons of a gene. Said introns may comprise further regulatory elements which are required for proper gene expression. Introns are usually transcribed together with the exons of a gene resulting in a nascent RNA transcript which contains both, exon and intron sequences. The intron encoded RNA sequences are usually removed by a process known as RNA splicing. However, said process also requires regulatory sequences present on a RNA transcript, said regulatory sequences may be encoded by the introns.

In addition, besides their function in transcriptional control and control of proper RNA processing and/or stability, regulatory elements of a gene could be also involved in the control of genetic stability of a gene locus. Said elements control, e.g., recombination events or serve to maintain a certain structure of the DNA or the arrangement of DNA in a chromosome.

Therefore, polymorphisms can occur in exons of a gene which encode an amino acid sequence as discussed supra as well as in regulatory regions which are involved in the above discussed process. The analysis of the nucleotide sequence of a gene locus in its entirety including, e.g., introns is in light of the above desirable. It has been found based on the polymorphisms comprised by the polynucleotides of the present invention that the mechanism of the increased expression of CYP3A5 protein in most Caucasians livers described in the examples below may involve enhanced transcription and stabilisation of the gene's transcripts.

Therefore, in a furthermore preferred embodiment of the gene of the invention a nucleotide deletion, addition and/or substitution results in altered expression of the variant gene compared to the corresponding wild type gene.

In another embodiment the present invention relates to a vector comprising the polynucleotide of the invention or the gene of the invention.

Said vector may be, for example, a phage, plasmid, viral or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host/cells.

The polynucleotides or genes of the invention may be joined to a vector containing selectable markers for propagation in a host. Generally, a plasmid vector is introduced in a precipitate such as a calcium phosphate precipitate, or in a complex with a charged lipid or in carbon-based clusters. Should the vector be a virus, it may be packaged in vitro using an appropriate packaging cell line prior to application to host cells.

In a more preferred embodiment of the vector of the invention the polynucleotide is operatively linked to expression control sequences allowing expression in prokaryotic or eukaryotic cells or isolated fractions thereof.

Expression of said polynucleotide comprises transcription of the polynucleotide, preferably into a translatable mRNA. Regulatory elements ensuring expression in eukaryotic cells, preferably mammalian cells, are well known to those skilled in the art. They usually comprise regulatory sequences ensuring initiation of transcription and optionally poly-A signals ensuring termination of transcription and stabilization of the transcript. Additional regulatory elements may include transcriptional as well as translational enhancers. Possible regulatory elements permitting expression in prokaryotic host cells comprise, e.g., the *lac*, *trp* or *tac* promoter in *E. coli*, and examples for regulatory elements permitting expression in eukaryotic host cells are the *AOX1* or *GAL1* promoter in yeast or the CMV-, SV40-, RSV-promoter (Rous sarcoma virus), CMV-enhancer, SV40-enhancer or a globin intron in mammalian and other animal cells. Beside elements which are responsible for the initiation of transcription such regulatory elements may also comprise transcription termination signals, such as the SV40-poly-A site or the tk-poly-A site, downstream of the polynucleotide. In this context, suitable expression vectors are known in the art such as Okayama-Berg cDNA expression vector pcDV1 (Pharmacia), pCDM8, pRc/CMV, pcDNA1, pcDNA3 (In-vitrogen), pSPORT1 (GIBCO BRL). Preferably, said vector is an expression vector and/or a gene transfer or targeting vector. Expression vectors derived from viruses such as retroviruses, vaccinia virus, adeno-associated virus, herpes viruses, or bovine

papilloma virus, may be used for delivery of the polynucleotides or vector of the invention into targeted cell population. Methods which are well known to those skilled in the art can be used to construct recombinant viral vectors; see, for example, the techniques described in Sambrook, Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory (1989) N.Y. and Ausubel, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y. (1994). Alternatively, the polynucleotides and vectors of the invention can be reconstituted into liposomes for delivery to target cells.

The term "isolated fractions thereof" refers to fractions of eukaryotic or prokaryotic cells or tissues comprising said cells which are capable of transcribing or transcribing and translating RNA from the vector of the invention. Said fractions comprise proteins which are required for transcription of RNA or transcription of RNA and translation of said RNA into a polypeptide. Said isolated fractions may be, e.g., nuclear and cytoplasmic fractions of eukaryotic cells such as of reticulocytes.

The present invention furthermore relates to a host cell genetically engineered with the polynucleotide of the invention, the gene of the invention or the vector of the invention.

Said host cell may be a prokaryotic or eukaryotic cell; see supra. The polynucleotide or vector of the invention which is present in the host cell may either be integrated into the genome of the host cell or it may be maintained extrachromosomally. In this respect, it is also to be understood that the recombinant DNA molecule of the invention can be used for "gene targeting" and/or "gene replacement", for restoring a mutant gene or for creating a mutant gene via homologous recombination; see for example Mouellic, Proc. Natl. Acad. Sci. USA, 87 (1990), 4712-4716; Joyner, Gene Targeting, A Practical Approach, Oxford University Press.

The host cell can be any prokaryotic or eukaryotic cell, such as a bacterial, insect, fungal, plant, animal or human cell. Preferred fungal cells are, for example, those of the genus *Saccharomyces*, in particular those of the species *S. cerevisiae*. The term "prokaryotic" is meant to include all bacteria which can be transformed or transfected with a polynucleotide for the expression of a variant polypeptide of the invention. Prokaryotic hosts may include gram negative as well as gram positive bacteria such as, for example, *E. coli*, *S. typhimurium*, *Serratia marcescens* and *Bacillus subtilis*. A polynucleotide coding for a mutant form of variant polypeptides of the invention can be used to transform or transfet the host using any of the techniques commonly known to those of ordinary skill in the art.

Methods for preparing fused, operably linked genes and expressing them in bacteria or animal cells are well-known in the art (Sambrook, *supra*). The genetic constructs and methods described therein can be utilized for expression of variant polypeptides of the invention in, e.g., prokaryotic hosts. In general, expression vectors containing promoter sequences which facilitate the efficient transcription of the inserted polynucleotide are used in connection with the host. The expression vector typically contains an origin of replication, a promoter, and a terminator, as well as specific genes which are capable of providing phenotypic selection of the transformed cells. The transformed prokaryotic hosts can be grown in fermentors and cultured according to techniques known in the art to achieve optimal cell growth. The proteins of the invention can then be isolated from the grown medium, cellular lysates, or cellular membrane fractions. The isolation and purification of the microbially or otherwise expressed polypeptides of the invention may be by any conventional means such as, for example, preparative chromatographic separations and immunological separations such as those involving the use of monoclonal or polyclonal antibodies.

Thus, in a further embodiment the invention relates to a method for producing a molecular variant polypeptide or fragment thereof comprising culturing the above described host cell; and recovering said protein or fragment from the culture.

In another embodiment the present invention relates to a method for producing cells capable of expressing a molecular variant polypeptide comprising genetically engineering cells with the polynucleotide of the invention, the gene of the invention or the vector of the invention.

The cells obtainable by the method of the invention can be used, for example, to test drugs according to the methods described in D. L. Spector, R. D. Goldman, L. A. Leinwand, *Cells*, a Lab manual, CSH Press 1998. Furthermore, the cells can be used to study known drugs and unknown derivatives thereof for their ability to complement the deficiency caused by mutations in the CYP3A5 gene. For these embodiments the host cells preferably lack a wild type allele, preferably both alleles of the CYP3A5 gene and/or have at least one mutated from thereof. Ideally, the gene comprising an allele as comprised by the polynucleotides of the invention could be introduced into the wild type locus by homologous replacement. Alternatively, strong overexpression of a mutated allele over the normal allele and comparison with a recombinant cell line overexpressing the normal

allele at a similar level may be used as a screening and analysis system. The cells obtainable by the above-described method may also be used for the screening methods referred to herein below.

Furthermore, the invention relates to a polypeptide or fragment thereof encoded by the polynucleotide of the invention, the gene of the invention or obtainable by the method described above or from cells produced by the method described above.

In this context it is also understood that the variant polypeptide of the invention can be further modified by conventional methods known in the art. By providing said variant proteins according to the present invention it is also possible to determine the portions relevant for their biological activity or inhibition of the same. The terms "polypeptide" and "protein" as used herein are exchangeable. Moreover, what is comprised by said terms is standard textbook knowledge.

The present invention furthermore relates to an antibody which binds specifically to the polypeptide of the invention.

Advantageously, the antibody specifically recognizes or binds an epitope containing one or more amino acid substitution(s) as defined above. Antibodies against the variant polypeptides of the invention can be prepared by well known methods using a purified protein according to the invention or a (synthetic) fragment derived therefrom as an antigen. Monoclonal antibodies can be prepared, for example, by the techniques as originally described in Köhler and Milstein, *Nature* 256 (1975), 495, and Galfré, *Meth. Enzymol.* 73 (1981), 3, which comprise the fusion of mouse myeloma cells to spleen cells derived from immunized mammals. In a preferred embodiment of the invention, said antibody is a monoclonal antibody, a polyclonal antibody, a single chain antibody, human or humanized antibody, primatized, chimerized or fragment thereof that specifically binds said peptide or polypeptide also including bispecific antibody, synthetic antibody, antibody fragment, such as Fab, Fv or scFv fragments etc., or a chemically modified derivative of any of these. Furthermore, antibodies or fragments thereof to the aforementioned polypeptides can be obtained by using methods which are described, e.g., in Harlow and Lane "Antibodies, A Laboratory Manual", CSH Press, Cold Spring Harbor, 1988. These antibodies can be used, for example, for the immunoprecipitation and immunolocalization of the variant polypeptides of the invention as well as for the monitoring of the presence of

said variant polypeptides, for example, in recombinant organisms, and for the identification of compounds interacting with the proteins according to the invention. For example, surface plasmon resonance as employed in the BIACore system can be used to increase the efficiency of phage antibodies which bind to an epitope of the protein of the invention (Schier, Human Antibodies Hybridomas 7 (1996), 97-105; Malmborg, J. Immunol. Methods 183 (1995), 7-13).

In a preferred embodiment the antibody of the present invention specifically recognizes an epitope containing one or more amino acid substitution(s) resulting from a nucleotide exchange as defined supra.

Antibodies which specifically recognize modified amino acids such as phospho-Tyrosine residues are well known in the art. Similarly, in accordance with the present invention antibodies which specifically recognize even a single amino acid exchange in an epitope may be generated by the well known methods described supra.

In light of the foregoing, in a more preferred embodiment the antibody of the present invention is monoclonal or polyclonal.

The invention also relates to a transgenic non-human animal comprising at least one polynucleotide of the invention, the gene of the invention or the vector of the invention as described supra.

The present invention also encompasses a method for the production of a transgenic non-human animal comprising introduction of a polynucleotide or vector of the invention into a germ cell, an embryonic cell, stem cell or an egg or a cell derived therefrom. The non-human animal can be used in accordance with the method of the invention described below and may be a non-transgenic healthy animal, or may have a disease or disorder, preferably a disease caused by at least one mutation in the gene of the invention. Such transgenic animals are well suited for, e.g., pharmacological studies of drugs in connection with variant forms of the above described variant polypeptides since these polypeptides or at least their functional domains are conserved between species in higher eukaryotes, particularly in mammals. Production of transgenic embryos and screening of those can be performed, e.g., as described by A. L. Joyner Ed., Gene Targeting, A Practical Approach

(1993), Oxford University Press. The DNA of the embryos can be analyzed using, e.g., Southern blots with an appropriate probe or based on PCR techniques.

A transgenic non-human animal in accordance with the invention may be a transgenic mouse, rat, hamster, dog, monkey, rabbit, pig, frog, nematode such as *Caenorhabditis elegans*, fruitfly such as *Drosophila melanogaster* or fish such as torpedo fish or zebrafish comprising a polynucleotide or vector of the invention or obtained by the method described above, preferably wherein said polynucleotide or vector is stably integrated into the genome of said non-human animal, preferably such that the presence of said polynucleotide or vector leads to the expression of the variant polypeptide of the invention. It may comprise one or several copies of the same or different polynucleotides or genes of the invention. This animal has numerous utilities, including as a research model for cancer or diseases including cardiovascular diseases, diabetes and AIDS or any other disease caused by as dysfunction or dysregulation of the polynucleotides or polypeptides of the invention research and therefore, presents a novel and valuable animal in the development of therapies, treatment, etc. for cancer diseases or diseases including cardiovascular diseases, diabetes and AIDS or any other disease caused by as dysfunction or dysregulation of the polynucleotides or polypeptides of the invention. Accordingly, in this instance, the mammal is preferably a laboratory animal such as a mouse or rat.

Thus, in a preferred embodiment the transgenic non-human animal of the invention is a mouse, a rat or a zebrafish.

Numerous reports revealed that said animals are particularly well suited as model organisms for the investigation of the drug metabolism and its deficiencies or cancer. Advantageously, transgenic animals can be easily created using said model organisms, due to the availability of various suitable techniques well known in the art.

The invention also relates to a solid support comprising one or a plurality of the polynucleotide, the gene, the vector, the polypeptide, the antibody or the host cell of the invention in immobilized form.

The term "solid support" as used herein refers to a flexible or non-flexible support that is suitable for carrying said immobilized targets. Said solid support may be homogenous or

inhomogeneous. For example, said solid support may consist of different materials having the same or different properties with respect to flexibility and immobilization, for instance, or said solid support may consist of one material exhibiting a plurality of properties also comprising flexibility and immobilization properties. Said solid support may comprise glass-, polypropylene- or silicon-chips, membranes, oligonucleotide-conjugated beads or bead arrays.

The term "immobilized" means that the molecular species of interest is fixed to a solid support, preferably covalently linked thereto. This covalent linkage can be achieved by different means depending on the molecular nature of the molecular species. Moreover, the molecular species may be also fixed on the solid support by electrostatic forces, hydrophobic or hydrophilic interactions or Van-der-Waals forces. The above described physico-chemical interactions typically occur in interactions between molecules. For example, biotinylated polypeptides may be fixed on a avidin-coated solid support due to interactions of the above described types. Further, polypeptides such as antibodies, may be fixed on an antibody coated solid support. Moreover, the immobilization is dependent on the chemical properties of the solid support. For example, the nucleic acid molecules can be immobilized on a membrane by standard techniques such as UV-crosslinking or heat.

In a preferred embodiment of the invention said solid support is a membrane, a glass- or polypropylene- or silicon-chip, are oligonucleotide-conjugated beads or a bead array, which is assembled on an optical filter substrate.

Moreover, the present invention relates to an in vitro method for identifying a polymorphism said method comprising the steps of:

- (a) isolating a polynucleotide or the gene of the invention from a plurality of subgroups of individuals, wherein one subgroup has no prevalence for a CYP3A5 associated disease and at least one or more further subgroup(s) do have prevalence for a CYP3A5 associated disease; and
- (b) identifying a polymorphism by comparing the nucleic acid sequence of said polynucleotide or said gene of said one subgroup having no prevalence for a CYP3A5 associated disease with said at least one or more further subgroup(s) having a prevalence for a CYP3A5 associated disease.

The term "prevalence" as used herein means that individuals are susceptible for one or more disease(s) which are associated with CYP3A5 dysfunction or dysregulation or could already have one or more of said disease(s). Thereby, one CYP3A5 associated disease can be used to determine the susceptibility for another CYP3A5 associated disease, e.g. impaired drug metabolism may be indicative for a prevalence for, e.g. cancer. Moreover, symptoms which are indicative for a prevalence for developing said diseases are very well known in the art and have been sufficiently described in standard textbooks such as Pschyrembel.

Advantageously, polymorphisms according to the present invention which are associated with CYP3A5 dysfunction or dysregulation or one or more disease(s) based thereon should be enriched in subgroups of individuals which have a prevalence for said diseases versus subgroups which have no prevalence for said diseases. Thus, the above described method allows the rapid and reliable detection of polymorphisms which are indicative for one or more CYP3A5 associated disease(s) or a susceptibility therefor. Advantageously, due to the phenotypic preselection a large number of individuals having no prevalence might be screened for polymorphisms in general. Thereby, a reference sequences comprising polymorphisms which do not correlate to one or more CYP3A5 associated disease(s) can be obtained. Based on said reference sequences it is possible to efficiently and reliably determine the relevant polymorphisms.

In a further embodiment the present invention relates to a method for identifying and obtaining a pro-drug or a drug capable of modulating the activity of a molecular variant of a CYP3A5 polypeptide comprising the steps of:

- (a) contacting the polypeptide, the solid support of the invention, a cell expressing a molecular variant gene comprising a polynucleotide of the invention, the gene or the vector of the invention in the presence of components capable of providing a detectable signal in response to drug activity with a compound to be screened for pro-drug or drug activity; and
- (b) detecting the presence or absence of a signal or increase or decrease of a signal generated from the pro-drug or the drug activity, wherein the absence, presence, increase or decrease of the signal is indicative for a putative pro-drug or drug.

The term "compound" in a method of the invention includes a single substance or a plurality of substances which may or may not be identical.

Said compound(s) may be chemically synthesized or produced via microbial fermentation but can also be comprised in, for example, samples, e.g., cell extracts from, e.g., plants, animals or microorganisms. Furthermore, said compounds may be known in the art but hitherto not known to be useful as an inhibitor, respectively. The plurality of compounds may be, e.g., added to the culture medium or injected into a cell or non-human animal of the invention.

If a sample containing (a) compound(s) is identified in the method of the invention, then it is either possible to isolate the compound from the original sample identified as containing the compound in question or one can further subdivide the original sample, for example, if it consists of a plurality of different compounds, so as to reduce the number of different substances per sample and repeat the method with the subdivisions of the original sample. It can then be determined whether said sample or compound displays the desired properties, for example, by the methods described herein or in the literature (Spector et al., Cells manual; see supra). Depending on the complexity of the samples, the steps described above can be performed several times, preferably until the sample identified according to the method of the invention only comprises a limited number of or only one substance(s). Preferably said sample comprises substances of similar chemical and/or physical properties, and most preferably said substances are identical. The methods of the present invention can be easily performed and designed by the person skilled in the art, for example in accordance with other cell based assays described in the prior art or by using and modifying the methods as described herein. Furthermore, the person skilled in the art will readily recognize which further compounds may be used in order to perform the methods of the invention, for example, enzymes, if necessary, that convert a certain compound into a precursor. Such adaptation of the method of the invention is well within the skill of the person skilled in the art and can be performed without undue experimentation.

Compounds which can be used in accordance with the present invention include peptides, proteins, nucleic acids, antibodies, small organic compounds, ligands, peptidomimetics, PNAs and the like. Said compounds may act as agonists or antagonists of the invention. Said compounds can also be functional derivatives or analogues of known drugs. Methods for the preparation of chemical derivatives and analogues are well known to those skilled in the art and are described in, for example, Beilstein, Handbook of Organic Chemistry, Springer edition New York Inc., 175 Fifth Avenue, New York, N.Y. 10010 U.S.A. and Organic Synthesis, Wiley, New York, USA. Furthermore, said derivatives and analogues

can be tested for their effects according to methods known in the art or as described. Furthermore, peptide mimetics and/or computer aided design of appropriate drug derivatives and analogues can be used, for example, according to the methods described below. Such analogs comprise molecules may have as the basis structure of known CYP3A5 substrates and/or inhibitors and/or modulators; see infra.

Appropriate computer programs can be used for the identification of interactive sites of a putative inhibitor and the polypeptides of the invention by computer assisted searches for complementary structural motifs (Fassina, Immunomethods 5 (1994), 114-120). Further appropriate computer systems for the computer aided design of protein and peptides are described in the prior art, for example, in Berry, Biochem. Soc. Trans. 22 (1994), 1033-1036; Wodak, Ann. N. Y. Acad. Sci. 501 (1987), 1-13; Pabo, Biochemistry 25 (1986), 5987-5991. The results obtained from the above-described computer analysis can be used in combination with the method of the invention for, e.g., optimizing known inhibitors, analogs, antagonists or agonists. Appropriate peptidomimetics and other inhibitors can also be identified by the synthesis of peptidomimetic combinatorial libraries through successive chemical modification and testing the resulting compounds, e.g., according to the methods described herein. Methods for the generation and use of peptidomimetic combinatorial libraries are described in the prior art, for example in Ostresh, Methods in Enzymology 267 (1996), 220-234 and Dorner, Bioorg. Med. Chem. 4 (1996), 709-715. Furthermore, the three-dimensional and/or crystallographic structure of said compounds and the polypeptides of the invention can be used for the design of peptidomimetic drugs (Rose, Biochemistry 35 (1996), 12933-12944; Rutenber, Bioorg. Med. Chem. 4 (1996), 1545-1558). It is very well known how to obtain said compounds, e.g. by chemical or biochemical standard techniques. Thus, also comprised by the method of the invention are means of making or producing said compounds. In summary, the present invention provides methods for identifying and obtaining compounds which can be used in specific doses for the treatment of specific forms of CYP3A5 associated diseases, e.g. dysfunctions of the drug metabolism or cancer.

The above definitions apply mutatis mutandis to all of the methods described in the following.

In a further embodiment the present invention relates to a method for identifying and obtaining an inhibitor of the activity of a molecular variant of a CYP3A5 polypeptide comprising the steps of:

- (a) contacting the protein, the solid support of the invention or a cell expressing a molecular variant gene comprising a polynucleotide or the gene or the vector of the invention in the presence of components capable of providing a detectable signal in response to drug activity with a compound to be screened for inhibiting activity; and
- (b) detecting the presence or absence of a signal or increase or decrease of a signal generated from the inhibiting activity, wherein the absence or decrease of the signal is indicative for a putative inhibitor.

In a preferred embodiment of the method of the invention said cell is a cell, obtained by the method of the invention or can be obtained from the transgenic non-human animal as described supra.

In a still further embodiment the present invention relates to a method of identifying and obtaining a pro-drug or drug capable of modulating the activity of a molecular variant of a CYP3A5 polypeptide comprising the steps of:

- (a) contacting the host cell, the cell obtained by the method of the invention, the polypeptide or the solid support of the invention with the first molecule known to be bound by a CYP3A5 polypeptide to form a first complex of said polypeptide and said first molecule;
- (b) contacting said first complex with a compound to be screened, and
- (c) measuring whether said compound displaces said first molecule from said first complex.

Advantageously, in said method said measuring step comprises measuring the formation of a second complex of said protein and said inhibitor candidate. Preferably, said measuring step comprises measuring the amount of said first molecule that is not bound to said protein.

In a particularly preferred embodiment of the above-described method of said first molecule is an agonist or antagonist or a substrate and/or an inhibitor and/or a modulator of the polypeptide of the invention, e.g., with a radioactive or fluorescent label.

In a still another embodiment the present invention relates to a method of identifying and obtaining an inhibitor capable of modulating the activity of a molecular variant of a CYP3A5 polypeptide comprising the steps of:

- (a) contacting the host cell or the cell obtained by the method of the invention, the protein or the solid support of the invention with the first molecule known to be bound by a CYP3A5 polypeptide to form a first complex of said protein and said first molecule;
- (b) contacting said first complex with a compound to be screened, and
- (c) measuring whether said compound displaces said first molecule from said first complex.

In a preferred embodiment of the method of the invention said measuring step comprises measuring the formation of a second complex of said protein and said compound.

In another preferred embodiment of the method of the invention said measuring step comprises measuring the amount of said first molecule that is not bound to said protein.

In a more preferred embodiment of the method of the invention said first molecule is labeled.

The invention furthermore relates to a method for the production of a pharmaceutical composition comprising the steps of the method as described supra; and the further step of formulating the compound identified and obtained or a derivative thereof in a pharmaceutically acceptable form.

The therapeutically useful compounds identified according to the methods of the invention can be formulated and administered to a patient as discussed above. For uses and therapeutic doses determined to be appropriate by one skilled in the art and for definitions of the term "pharmaceutical composition" see infra.

Furthermore, the present invention encompasses a method for the preparation of a pharmaceutical composition comprising the steps of the above-described methods; and formulating a drug or pro-drug in the form suitable for therapeutic application and preventing or ameliorating the disorder of the subject diagnosed in the method of the invention.

Drugs or pro-drugs after their *in vivo* administration are metabolized in order to be eliminated either by excretion or by metabolism to one or more active or inactive

metabolites (Meyer, J. Pharmacokinet. Biopharm. 24 (1996), 449-459). Thus, rather than using the actual compound or inhibitor identified and obtained in accordance with the methods of the present invention a corresponding formulation as a pro-drug can be used which is converted into its active in the patient. Precautionary measures that may be taken for the application of pro-drugs and drugs are described in the literature; see, for review, Ozama, J. Toxicol. Sci. 21 (1996), 323-329).

In a preferred embodiment of the method of the present invention said drug or prodrug is a derivative of a medicament as defined hereinafter.

The present invention also relates to a method of diagnosing a disorder related to the presence of a molecular variant of the CYP3A5 gene or susceptibility to such a disorder comprising determining the presence of a polynucleotide or the gene of the invention in a sample from a subject.

In accordance with this embodiment of the present invention, the method of testing the status of a disorder or susceptibility to such a disorder can be effected by using a polynucleotide gene or nucleic acid of the invention, e.g., in the form of a Southern or Northern blot or *in situ* analysis. Said nucleic acid sequence may hybridize to a coding region of either of the genes or to a non-coding region, e.g. intron. In the case that a complementary sequence is employed in the method of the invention, said nucleic acid molecule can again be used in Northern blots. Additionally, said testing can be done in conjunction with an actual blocking, e.g., of the transcription of the gene and thus is expected to have therapeutic relevance. Furthermore, a primer or oligonucleotide can also be used for hybridizing to one of the above mentioned CYP3A5 gene or corresponding mRNAs. The nucleic acids used for hybridization can, of course, be conveniently labeled by incorporating or attaching, e.g., a radioactive or other marker. Such markers are well known in the art. The labeling of said nucleic acid molecules can be effected by conventional methods.

Additionally, the presence or expression of variant CYP3A5 gene can be monitored by using a primer pair that specifically hybridizes to either of the corresponding nucleic acid sequences and by carrying out a PCR reaction according to standard procedures. Specific hybridization of the above mentioned probes or primers preferably occurs at stringent hybridization conditions. The term "stringent hybridization conditions" is well known in the art; see, for example, Sambrook et al., "Molecular Cloning, A Laboratory Manual" second

ed., CSH Press, Cold Spring Harbor, 1989; "Nucleic Acid Hybridisation, A Practical Approach", Hames and Higgins eds., IRL Press, Oxford, 1985. Furthermore, the mRNA, cRNA, cDNA or genomic DNA obtained from the subject may be sequenced to identify mutations which may be characteristic fingerprints of mutations in the polynucleotide or the gene of the invention. The present invention further comprises methods wherein such a fingerprint may be generated by RFLPs of DNA or RNA obtained from the subject, optionally the DNA or RNA may be amplified prior to analysis, the methods of which are well known in the art. RNA fingerprints may be performed by, for example, digesting an RNA sample obtained from the subject with a suitable RNA-Enzyme, for example RNase T<sub>1</sub>, RNase T<sub>2</sub> or the like or a ribozyme and, for example, electrophoretically separating and detecting the RNA fragments as described above.

Further modifications of the above-mentioned embodiment of the invention can be easily devised by the person skilled in the art, without any undue experimentation from this disclosure; see, e.g., the examples. An additional embodiment of the present invention relates to a method wherein said determination is effected by employing an antibody of the invention or fragment thereof. The antibody used in the method of the invention may be labeled with detectable tags such as a histidine flags or a biotin molecule.

The invention relates to a method of diagnosing a disorder related to the presence of a molecular variant of a CYP3A5 gene or susceptibility to such a disorder comprising determining the presence of a polypeptide or the antibody of the invention in a sample from a subject.

In a preferred embodiment of the above described method said disorder is cancer or diseases including cardiovascular diseases, diabetes and AIDS.

In a preferred embodiment of the present invention, the above described method is comprising PCR, ligase chain reaction, restriction digestion, direct sequencing, nucleic acid amplification techniques, hybridization techniques or immunoassays.

Said techniques are very well known in the art.

Moreover, the invention relates to a method of detection of the polynucleotide or the gene of the invention in a sample comprising the steps of

- (a) contacting the solid support described supra with the sample under conditions allowing interaction of the polynucleotide or the gene of the invention with the immobilized targets on a solid support and;
- (b) determining the binding of said polynucleotide or said gene to said immobilized targets on a solid support.

The invention also relates to an in vitro method for diagnosing a disease comprising the steps of the method described supra, wherein binding of said polynucleotide or gene to said immobilized targets on said solid support is indicative for the presence or the absence of said disease or a prevalence for said disease.

The invention furthermore relates to a diagnostic composition comprising the polynucleotide, the gene, the vector, the polypeptide or the antibody of the invention.

In addition, the invention relates to a pharmaceutical composition comprising the polynucleotide, the gene, the vector, the polypeptide or the antibody of the invention.

These pharmaceutical compositions comprising, e.g., the antibody may conveniently be administered by any of the routes conventionally used for drug administration, for instance, orally, topically, parenterally or by inhalation. Acceptable salts comprise acetate, methylester, HCl, sulfate, chloride and the like. The compounds may be administered in conventional dosage forms prepared by combining the drugs with standard pharmaceutical carriers according to conventional procedures. These procedures may involve mixing, granulating and compressing or dissolving the ingredients as appropriate to the desired preparation. It will be appreciated that the form and character of the pharmaceutically acceptable character or diluent is dictated by the amount of active ingredient with which it is to be combined, the route of administration and other well-known variables. The carrier(s) must be "acceptable" in the sense of being compatible with the other ingredients of the formulation and not deleterious to the recipient thereof. The pharmaceutical carrier employed may be, for example, either a solid or liquid. Exemplary of solid carriers are lactose, terra alba, sucrose, talc, gelatin, agar, pectin, acacia, magnesium stearate, stearic acid and the like. Exemplary of liquid carriers are phosphate buffered saline solution, syrup, oil such as peanut oil and olive oil, water, emulsions, various types of wetting agents, sterile solutions and the like. Similarly, the carrier or diluent may include time delay

material well known to the art, such as glyceryl mono-stearate or glyceryl distearate alone or with a wax.

The dosage regimen will be determined by the attending physician and other clinical factors; preferably in accordance with any one of the above described methods. As is well known in the medical arts, dosages for any one patient depends upon many factors, including the patient's size, body surface area, age, the particular compound to be administered, sex, time and route of administration, general health, and other drugs being administered concurrently. Progress can be monitored by periodic assessment.

Furthermore, the use of pharmaceutical compositions which comprise antisense-oligonucleotides which specifically hybridize to RNA encoding mutated versions of the polynucleotide or gene according to the invention or which comprise antibodies specifically recognizing a mutated polypeptide of the invention but not or not substantially the functional wild-type form is conceivable in cases in which the concentration of the mutated form in the cells should be reduced.

Thanks to the present invention the particular drug selection, dosage regimen and corresponding patients to be treated can be determined in accordance with the present invention. The dosing recommendations will be indicated in product labeling by allowing the prescriber to anticipate dose adjustments depending on the considered patient group, with information that avoids prescribing the wrong drug to the wrong patients at the wrong dose.

In another embodiment the present invention relates to the use of the polynucleotide, a polynucleotide comprising SEQ ID No: 104, a polynucleotide encoding a polypeptide comprising SEQ ID No: 145, the gene, the vector or the polypeptide of the invention, a polypeptide comprising SEQ ID No: 145 or the antibody of the invention for the preparation of a diagnostic composition for diagnosing a disease.

A gene encoding a functional and expressible polypeptide of the invention can be introduced into the cells which in turn produce the protein of interest. Gene therapy, which is based on introducing therapeutic genes into cells by *ex-vivo* or *in-vivo* techniques is one of the most important applications of gene transfer. Suitable vectors and methods for *in-vitro* or *in-vivo* gene therapy are described in the literature and are known to the person skilled in the art; see, e.g., Giordano, Nature Medicine 2 (1996), 534-539; Schaper, Circ. Res. 79 (1996), 911-919; Anderson, Science 256 (1992), 808-813; Isner, Lancet 348 (1996), 370-374; Muhlhauser, Circ. Res. 77 (1995), 1077-1086; Wang, Nature Medicine 2

(1996), 714-716; WO94/29469; WO 97/00957 or Schaper, Current Opinion in Biotechnology 7 (1996), 635-640, and references cited therein. The gene may be designed for direct introduction or for introduction via liposomes, or viral vectors (e.g. adenoviral, retroviral) into the cell. Preferably, said cell is a germ line cell, embryonic cell, or egg cell or derived therefrom, most preferably said cell is a stem cell. As is evident from the above, it is preferred that in the use of the invention the nucleic acid sequence is operatively linked to regulatory elements allowing for the expression and/or targeting of the polypeptides of the invention to specific cells. Suitable gene delivery systems that can be employed in accordance with the invention may include liposomes, receptor-mediated delivery systems, naked DNA, and viral vectors such as herpes viruses, retroviruses, adenoviruses, and adeno-associated viruses, among others. Delivery of nucleic acids to a specific site in the body for gene therapy may also be accomplished using a biostatic delivery system, such as that described by Williams (Proc. Natl. Acad. Sci. USA 88 (1991), 2726-2729). Standard methods for transfecting cells with recombinant DNA are well known to those skilled in the art of molecular biology, see, e.g., WO 94/29469; see also supra. Gene therapy may be carried out by directly administering the recombinant DNA molecule or vector of the invention to a patient or by transfecting cells with the polynucleotide or vector of the invention ex vivo and infusing the transfected cells into the patient.

A polynucleotide comprising SEQ ID No: 104 and a polypeptide comprising SEQ ID No: 145 have already been described in Jounaidi *et al.* (Jounaidi, Biochem Biophys Res Commun 221 (1996), 466-70). However, Jounaidi *et al.* have merely disclosed the respective amino acid and nucleotide sequences without making any suggestion towards the pharmaceutic and diagnostic value of said polynucleotide or polypeptide, in particular for those disorders and diseases referred to infra.

In a further embodiment the present invention relates to the use of the polynucleotide, a polynucleotide comprising SEQ ID No: 104, a polynucleotide encoding a polypeptide comprising SEQ ID No: 145, the gene, the vector, the polypeptide of the invention, a polypeptide comprising SEQ ID No.: 145 or the antibody of the invention for the preparation of a pharmaceutical composition for treating a disease.

In another embodiment the present invention encompasses the use of a polynucleotide selected from the group consisting of:

- (a) a polynucleotide having the nucleic acid sequence of SEQ ID NO: 082, 088, 104, 112, 126, 131, or 140;

- (b) a polynucleotide encoding a polypeptide having the amino acid sequence of SEQ ID No: 127, 132, 141 or 145;
- (c) a polynucleotide capable of hybridizing to a CYP3A5 gene, wherein said polynucleotide is having at least one additional nucleotide at a position corresponding to position 3709/3710 or 27131/27132 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614) or a nucleotide exchange at a position corresponding to position 7303 or 27289 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614);
- (d) a polynucleotide capable of hybridizing to a CYP3A5 gene, wherein said polynucleotide is having an additional G nucleotide at a position corresponding to position 3709/3710 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614), an additional T nucleotide at a position corresponding to position 27131/27132 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614), or an A at a position corresponding to position 7303 or 27289 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614);

for the preparation of a diagnostic composition for diagnosing a disease in a subject having a genome comprising a variant allele of the CYP3A5 gene, wherein said allele is having an A at a position corresponding to position 6986 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614).

The definitions of the terms referred to in this specification apply mutatis mutandis to the aforementioned use.

The term "subject" inter alia refers to animals. Preferably, said animals belong to the animal species referred to above. Moreover, the term "subject" encompasses humans. The humans in accordance with the use of the present invention are selected from all existing ethnical groups and subgroups, e.g. Caucasians, African Americans or Asians. However, particular well suited for the use of the invention are African Americans for which it could be demonstrated that diagnosing a disease or a prevalence for a disease based on monitoring the presence or absence of an CYP3A5 allele having at a position corresponding to position 6986 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614) an A results in a false positive prediction for CYP3A5 expression in a considerable number of subjects. This allele of CYP3A5 has been described in detail in Kuehl, 2001, Nature Genetics 27: 383-391 as CYP3A5\*1 allele. The CYP3A5\*1 allele is characterized by the presence of an A at position 22893 of the CYP3A5 nucleic acid sequence referred to in Kuehl, loc.cit. The allelic frequency of said allele is particularly high in African Americans although it is also present in other ethnical groups or subgroups. However, it was found in accordance with the present invention that the CYP3A5 expression of subjects for which a false positive result was obtained in diagnostic studies based on the CYP3A5\*1 allele could be correctly predicted by further diagnosing the presence or absence of a polynucleotide as defined under (a) to (d) in accordance with the use of the present invention. For example, a polynucleotide having an additional nucleotide at a position corresponding to position 27131/27132 of the CYP3A5 gene as defined supra has been found in accordance with this invention to be present in approximately 10% of the African Americans resulting in a frameshift mutation in exon 11. The present invention provides means and methods to distinguish between the haplotype resulting in improved expression of CYP3A5 comprising the polymorphism(s) of the CYP3A5\*1 allele and the haplotype resulting in decreased expression, wherein said haplotype as set forth above comprises in addition to the polymorphism(s) of the CYP3A5\*1 allele co-segregating polymorphisms comprised by a polynucleotide referred to under (a) to (d) supra. Thus, based on the aforementioned use of the present invention a reliable diagnosis of the CYP3A5 activity of a subject is achieved.

The invention also relates to the use of a polynucleotide comprising a polynucleotide having an A at a position corresponding to position 14690 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614) for the preparation of a diagnostic composition for diagnosing a disease in a subject having a genome comprising a variant allele of the CYP3A5 gene, wherein said allele is having an A at a position corresponding to position 6986 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614).

The definitions of the terms referred to in this specification apply mutatis mutandis to the aforementioned use.

In accordance with the present invention it could also be shown that the polymorphism(s) constituting the CYP3A5\*1 and the CYP3A5\*6 allele as described in Kuehl, loc.cit., cosegregate in a considerable number of subjects and thereby constitute another haplotype resulting in decreased CYP3A5 expression. It has been shown that the CYP3A5\*6 allele results in inappropriate splicing of exon 7 of CYP3A5, a frameshift and a premature termination at position 184 of the CYP3A5 protein. Consequently, false positive result as regards the expression level of CYP3A5 in a subject can be obtained in diagnostic studies based on the CYP3A5\*1 allele. Said false positive results, however, can be avoided according to the use of this invention by further diagnosing the presence or absence of the polymorphism(s) of the CYP3A5\*6 allele. Thus, based on the aforementioned use of the present invention a reliable diagnosis of the CYP3A5 activity of a subject is achieved.

In light with the foregoing, in a preferred embodiment of the aforementioned use said subject is an African American.

As has been discussed above, the number of subjects which are diagnosed false positive is due to the high allelic frequency of CYP3A5 alleles such as those comprising a polynucleotide as defined under (a) to (d) above resulting in a frame shift mutation or those comprised by the CYP3A5\*6 allele. Said allelic frequency is particularly high within the group of African Americans. In accordance with the present invention it has been found that the CYP3A5\*6 allele is present in about 13.3% of the African Americans. Thus within this ethnic group the problems emerging from a wrong prediction of CYP3A5 expression are more severe than for other ethnical groups.

In a more preferred embodiment of the use of the present invention said disease is cancer or diseases including cardiovascular diseases, diabetes and AIDS.

Finally, the present invention relates to a diagnostic kit for detection of a polymorphism comprising the polynucleotide, the gene, the vector, the polypeptide, the antibody, the host cell, the transgenic non-human animal or the solid support of the invention.

The kit of the invention may contain further ingredients such as selection markers and components for selective media suitable for the generation of transgenic cells and animals. The kit of the invention can be used for carrying out a method of the invention and could be, *inter alia*, employed in a variety of applications, e.g., in the diagnostic field or as research tool. The parts of the kit of the invention can be packaged individually in vials or in combination in suitable containers or multicontainer units. Manufacture of the kit follows preferably standard procedures which are known to the person skilled in the art. The kit may be used for methods for detecting expression of a mutant form of the polypeptides, genes or polynucleotides in accordance with any one of the above-described methods of the invention, employing, for example, immunoassay techniques such as radioimmunoassay or enzymeimmunoassay or preferably nucleic acid hybridization and/or amplification techniques such as those described herein before and in the Examples as well as pharmacokinetic studies when using non-human transgenic animals of the invention.

The figures illustrate the invention.

**Figure 1:** **A.** Western blot analysis of CYP3A5 protein expression in microsomes prepared from 6 LE (low expressing) and 6 HE (high expressing) Caucasian livers. **B.** The relative contributions of CYP3A5 and CYP3A4 to the combined CYP3A5/CYP3A4 protein pool in 17 HE livers as determined by Western blot.

**Figure 2:** Expression levels and the allelic source of *CYP3A5* transcripts in LE and HE Caucasian livers. **A.** TaqMan analysis of *CYP3A5* mRNA in 8 LE (low expressing, white bars) and 8 HE (high expressing, grey bars) liver samples. **B.** Sequences of a portion of the 3'-UTR of *CYP3A5* in samples heterozygous for variant ch-v-015 (Table 2A). Templates used for PCR were genomic DNA (left panel), cDNA from a LE liver (middle panel) and cDNA from an HE liver (right panel).

**Figure 3:** Allelic frequencies of *CYP3A5* genetic variants in Caucasians. **A.** in DNA samples derived from 8-168 LE individuals **B.** in DNA samples derived from 7-18 HE individuals. The bars at the bottom of the figure indicate schematically the localisation of the pseudoexons PS2 exon 1 and 2 and exons 1 – 13 of the *CYP3A5* gene. The arrowhead marks the duplication boundary (Gellner, Pharmacogenetics 11 (2001), 111 - 121).

**Figure 4:** Genomic and peptide sequences: genomic DNA sequences containing the amplified regions in which polymorphisms were detected and polypeptide sequences with amino acid substitutions. Nucleotide sequences are listed in 5' - 3' orientation. Letters in lowercase indicate non-coding sequences, letters in uppercase indicate coding sequences. Primer regions are underlined. Variant sites are shown framed. Peptide sequences are shown in one letter code. || marks a site where a deletion has occurred. In Seq ID 198, the hybridizing site of the TaqMan® probe is shown in bold.

**Figure 5:** CYP3A5 cDNA insert region of the plasmid that was used as starting material for in vitro mutagenesis. Cloning sites are shown underlined. Modified 5' and 3' regions of the CYP3A5 cDNA are shown in lowercase letters. The 5' modification, a MALLAVF amino acid sequence on protein level, has been introduced in order to increase expression in *E. coli* (Gillam, Arch Biochem Biophys 317 (1995), 374-84). The 3' modification, a His<sub>6</sub> tag on protein level, has been introduced in order to enhance subsequent purification. The unmodified part of the CYP3A5 insert was verified to be identical to the CYP3A5 cDNA corresponding to accession no. NM\_000777.1 by sequencing, which is the underlying nucleotide sequence for NP\_000768.1. Sites corresponding to ch-v-009 and ch-v-001 are shown framed.

The invention will now be described by reference to the following biological Examples which are merely illustrative and are not construed as a limitation of the scope of the present invention.

**Example 1: Isolation of genomic DNA from human blood, generation and purification of CYP3A5 gene fragments**

Genomic DNA was isolated from blood or liver samples using Qiagen blood and tissue DNA isolation kits. Oligonucleotides used in the screen were designed based on the recently determined sequence and organisation of the human CYP3A locus (Gellner, Pharmacogenetics 11 (2001), 111 - 121). Primer sequences and PCR fragment lengths are given in Table 1A. Amplified fragments were processed through PCR purification columns (Qiagen) and sequenced on PE ABI 3700 DNA Analysers using the same primers as in PCR. The sequences were analysed for the presence of polymorphisms using the PHRED/PHRAP/POLYPHRED/CONSED software package (University of Washington, Seattle, WA, USA).

Total RNA was isolated from liver samples using the RNeasy kit (Qiagen) according to the manufacturers instructions except that an additional DNase I digestion was performed directly on the column. cDNA pools were generated from 1 µg of total RNA using random hexamer primers and Superscript reverse transcriptase (Life Technologies). The cDNA used for one TaqMan assay was derived from 40 ng total RNA. CYP3A5 mRNA expression levels were quantified by real time quantitative PCR using the ABI 7700

Sequence Detection System (PE Biosystems). Oligonucleotides and probes were designed with the Primer Express (PE Biosystems) programme. Oligonucleotides used for the quantitative PCR were: forward 5'- TTG TTG GGA AAT GTT TTG TCC TAT C -3' (Seq ID: 237) and reverse 5'- ACA GGG AGT TGA CCT TCA TAC GTT -3' (Seq ID: 238). The TaqMan probe (5'- TCA GGG TCT CTG GAA ATT TGA CAC AGA GTG CTA-3'; Seq ID: 239) was labelled with the 5' reporter dye 6-carboxy-fluorescein (FAM) and the 3' quencher 6-carboxy-tetramethylrhodamin (TAMRA). The experiments were performed according to a standard protocol developed by PE Biosystems. The specificity of the assay for *CYP3A5* was determined using equal amounts of *CYP3A4*, *CYP3A5*, *CYP3A7* and *CYP3A43* cDNA species expressed *in vitro*. The specificity of the probe was 10<sup>4</sup> times higher for *CYP3A5* than for *CYP3A7* cDNA whereas *CYP3A4* and *CYP3A43* cDNAs were not detectable at all. The linear range of the *CYP3A5* assay was determined to be between 10 and 10<sup>6</sup> target molecules. *CYP3A5* expression levels were normalised using the expression of 18S mRNA species determined with pre-developed TaqMan assays (PE Biosystems).

#### **Example 2: Determination of genetic variations within the *CYP3A5* locus**

Sequence diversity within the *CYP3A5* locus was determined by PCR amplification from genomic DNA (fragment size: 264 – 997 bp) and sequencing each PCR-product of 19 - 217 samples of Caucasian origin, 36 - 45 samples of African American origin, 34 - 47 samples of Chinese origin, 41 - 50 samples of Japanese origin, and 31 - 47 samples of Korean origin. The PCR-fragments encompass the entire protein-coding region of *CYP3A5*, a portion of the 3'-UTR, the entire 5'-UTR as well as 6203 bp sequence between the *CYP3A5* transcriptional start site and a L1\_5'UTR\_ORF repeat located upstream of the gene (Fig. 3). In addition, we genotyped two linked single nucleotide polymorphisms (SNPs, ch-v-020, ch-v-021, Table 2A-E) located in a sequence originally described as *CYP3A5* promoter that were recently reported to co-segregate with increased *CYP3A5* protein expression (Paulussen, Pharmacogenetics 10 (2000), 415-24). The results also indicate co-segregation of both variants. Using the recently determined sequence of the entire *CYP3A* locus (Gellner, Pharmacogenetics 11 (2001), 111 - 121), we place these variants approximately 20 kb upstream of the first exon of *CYP3A5*, in a sequence 5' adjacent to a *CYP3A* pseudogene (PS2 in Fig. 3). Furthermore, we additionally genotyped a single nucleotide polymorphism located in intron 3 of the *CYP3A5* gene (ch-v-048;

Kuehl, 2001, Nature Genetics 27: 383-391) by TaqMan® assay using the primers and probes listed in Table 1B. The analysis were performed on a Sequence Detection System (PE Biosystems).

A total of 29 variants including the two linked SNPs described by Paulussen *et al.* (Paulussen, Pharmacogenetics 10 (2000), 415-24) were detected in the screen of Caucasian samples and their allelic frequencies were estimated to be between 0.3 % and 11.9 % (Table 2A). 6 variants are located within the 6 kb sequence upstream of the transcriptional start site of *CYP3A5*. 14 variants are located in introns, or in the 5'-UTR or 3'-UTR, whereas 4 have been found in the protein-coding sequence. Among the latter ones, three variants result in amino acid substitutions and one in a premature termination of the *CYP3A5* protein (Table 2A). The g.7303C>A variant (ch-v-009, Table 2A) results in a S100Y amino acid exchange in exon 4. The g.3705C>T variant (ch-v-005) leads to a H30Y amino acid exchange in exon 2. Cloning and sequencing revealed a physical linkage of this variant to the g.3709-3710insG variant (ch-v-006). The latter variant results in a shift of the open reading frame leading to a truncation of the protein sequence at position 34 (K34.). The T398N variant (ch-v-001, Table 2A), originally described by Jounaidi (Jounaidi, Biochem Biophys Res Commun 221 (1996), 466-70), was found in 3 out of 80 individuals tested.

Neither of the four protein altering variants found in Caucasians have been found in the African-Americans, Chinese, Japanese or Korean samples (Table 2A-E). However, among others we have found 4 new variants in these samples that result in an altered *CYP3A5* amino acid sequence (ch-v-017, ch-v-043, ch-v-045, ch-v-068, Table 2B-E). The g.27131-27132insT (ch-v-017) variant in exon 11 (ch-v-017, Table 2B, 2D) has been found in 9 out of 45 African-American samples and in one out of 50 Japanese samples. The variant results in a shift of the open reading frame which leads to a truncation of the protein sequence at position 348 (D348.). Variants ch-v-043, ch-v-045 and ch-v-068 lead to amino acid exchanges.

### **Example 3: Identification of genetic determinants of CYP3A5 protein expression**

In the following, the frequencies of Caucasian *CYP3A5* gene variants have been analyzed as a function of *CYP3A5* protein expression. For this purpose, allelic frequencies of

variants shown in Table 2A were calculated separately for HE and LE livers (Fig. 3). The frequencies of 9 variants (ch-v-020, ch-v-021, ch-v-026, ch-v-034, ch-v-007, ch-v-008, ch-v-011, ch-v-014 and ch-v-015) were significantly increased in HE livers (all  $\chi^2 > 13.3$ , df = 1, p < 0.01, Bonferroni corrected). Except one, all tested HE livers (17/18, 94 %) were heterozygous for three variants (ch-v-021, ch-v-026 and ch-v-015). 16 of those samples were heterozygous for ch-v-020 as well. One HE sample could not be genotyped for this variant. In contrast, LE livers were either wildtype (155/168, 92.3 %), heterozygous for variants ch-v-021 and ch-v-26 (9/168, 5.4 %) or heterozygous for the variant ch-v-015 (4/168, 2.4 %) only. However, in LE livers all three variants never occurred simultaneously (Table 3). These results defined either of the three variants as a useful but imperfect marker of increased CYP3A5 expression. The variants ch-v-034, ch-v-008, ch-v-011 and ch-v-14 only occurred in a subset of the samples heterozygous for the above three variants (ch-v-021, ch-v-026 and ch-v-015).

The distribution of variants ch-v-021, ch-v-026 and ch-v-015 in the samples screened strongly suggest that they constitute a haplotype. In the following, the hypotheses whether these three variants recombine independently or not has been tested. Assuming their independent inheritance, the expected 3-loci-genotype frequencies for all combinations of variants and compared them with the observed frequencies have been calculated. The difference is highly significant ( $\chi^2 = 93.6$ ; classes 'all wildtype', 'single variant hetero- or homozygous', 'two or three variants hetero- or homozygous'; df = 1; p << 0.001). There were more individuals with two or three of the variants than expected and less individuals with only one of the variants. This result suggests linkage among the three variants. The degree of linkage with the linkage disequilibrium parameter D for the three pairs of variants was estimated. Using maximum likelihood estimates for haplotype frequencies, D was calculated to be 0.041 for the variant pairs ch-v-021/ch-v-015 and ch-v-026/ch-v-015, which is 80 % of its theoretical maximum, and 0.065 for variants ch-v-021 and ch-v-026 which corresponds to 100 % of its theoretical maximum.

The probability that individuals showing the respective variant genotype are HE (positive predictive value) is estimated to be 65 % for variants ch-v-021 and ch-v-026, respectively, and 81 % for the ch-v-015 variant. For the combination of all three variants the positive predictive value is 100 % in our sample set. However, assuming that these variants need to be located in *cis* for increased protein expression, it is clear that there is some probability for individuals showing all three variants to be LE. The results show that at least the allele ch-v-021/ch-v-026 and the allele ch-v-015 actually exist (see genotype 2 and 3,

Table 3) and therefore the existence of a genotype with a combination of these two alleles has to be postulated. The maximum likelihood estimate for the frequency of these 3-fold heterozygotes having not all three variants in *cis* is 0.05 % of all samples screened or 0.61 % of samples hetero- or homozygote for all three variants. In other words, of 100 Caucasians screened statistically about 9 of them will be hetero- or homozygous for all three variants and about 0.05 of these will have not all three variants in *cis*. Therefore, it can be expected that the positive predictive power of the 3-variant genotype to be about 99.95 %. Of course, the same values would be achieved for a combination of only two variants, either ch-v-021/ch-v-015 or ch-v-26/ch-v-15.

In a single HE liver none of the above 9 variants that were found in the other 17 HE samples could be detected. A closer examination of variants found in this sample revealed a variant within intron 4 (ch-v-018) and one within intron 5 (ch-v-019), respectively. These variants were unique to this sample, since they were not found in any other of the samples screened. Neither were they found in any of the other ethnic groups screened. It remains to be shown whether these variants are themselves causative for transcriptional activation or whether they are linked to another, so far undetected variant.

#### **Example 4: Determination of CYP3A5 protein expression**

Protein expression of CYP3A4 and CYP3A5 in Caucasian liver samples was determined by Western blotting using CYP3A4- and CYP3A5-specific antibodies (Gentest). Liver microsomes were prepared as previously described (Zanger, Biochemistry 27 (1988), 5447-54). To obtain total protein homogenate, powdered liver tissue was homogenised in 0.1 M Tris-Cl pH 7.4, 1 mM EDTA, 1 mM Pefabloc SC, 1 µg/ml leupeptin, 1 µg/ml pepstatin with a Potter Elvehjem homogenisator (glass/Teflon) for 2 min at 1000 rpm. Homogenates were then sonified with a Bandelin Sonoplus HD 200 and stored at -80°C. For Western blotting, 12.5 µg microsomal protein homogenate or 40 µg total protein homogenate were separated in a 10 % SDS-polyacrylamide gel. Electrophoretic transfer onto PVDF membranes was carried out in a TankBlot Cell (BioRad) for 1.5 hours at constant voltage (100 V) and at 10 °C. Following the transfer, the membranes were incubated for 60 min in 5 % milk, TBS, 0.1 % Tween 20 to reduce the unspecific antibody binding. Incubations with either primary antibody (Gentest, dilution 1:500) were performed in 1 % milk, TBS, 0.1 % Tween 20 for 60 min, those with the secondary antibody (anti-

rabbit IgG-POD Fab-fragments, Dianova, dilution 1:10000 in the same solution for 30 min. CYP3A4 or CYP3A5 protein bands were detected with Supersignal Dura (Pierce) and a digital CCD-camera (LAS-1000, Fuji). Signal quantification was performed with AIDA (Raytest). Protein expression levels were calculated based on calibration curves obtained with microsomes expressing recombinant CYP3A4 and CYP3A5 proteins (Gentest).

Homogenates or microsomal fractions were prepared from 186 human livers and investigated by Western blotting using a CYP3A5-specific antibody. CYP3A5 protein was detected in all samples analysed and its expression showed a clear bimodality (Fig. 1A). 168 livers (~ 90 %), further referred to as LE (low-expressing), showed expression close to or below the lower limit of quantification (LLOQ) of the assay (0.3 pmol/mg homogenate protein and 1.0 pmol/mg microsomal protein). Eighteen samples (~ 10 %), further referred to as HE (high-expressing), exhibited much higher CYP3A5 expression levels. The expression was in the range between 1.6 and 2.9 pmol/mg homogenate protein ( $2.3 \pm 0.5$ ; n = 6) and between 3.9 and 15.5 pmol/mg microsomal protein ( $9.7 \pm 4.1$ ; n = 12). Taking the LLOQ of the assay as the expression level of CYP3A5 in LE livers, HE livers express on average 8 to 10 times more CYP3A5 protein than LE livers.

In the following, the contribution of CYP3A5 to the combined CYP3A5 and CYP3A4 protein expression in HE livers was investigated. CYP3A4 expression in these livers was between 0.9 and 82.6 pmol/mg homogenate protein (n = 6) and between 4.5 and 295 pmol/mg microsomal protein (n = 11). The levels and the range of CYP3A4 variability in HE livers were similar to those in LE livers (not shown). Fig. 1B shows the share of CYP3A5 in the combined CYP3A5 and CYP3A4 protein pool in 17 HE livers. CYP3A5 contribution varies between 3 % and 74 %. In an average HE liver, the share of CYP3A5 in the combined pool of both proteins is 24 %. Taking the LLOQ of the CYP3A5 assay as the actual expression level of the protein in LE livers, the corresponding value in these livers is approximately 1.6 %.

#### Example 5: Determination of CYP3A5 mRNA expression

The expression of CYP3A5 mRNA in 8 Caucasian HE and 8 LE livers was investigated using a CYP3A5-specific TaqMan probe. As illustrated in Fig. 2A, the distribution of CYP3A5 mRNA levels exhibited a bimodality which was in complete agreement with that

observed in the expression of CYP3A5 protein (Fisher's exact test,  $p = << 0.001$ ). The number of 3A5 transcripts per ng of total RNA in HE livers ( $n = 8$ ) was on average 8.5 times higher than those in LE livers ( $n = 8$ ).

In the following, the allelic origin of *CYP3A5* transcripts in HE and LE livers was investigated. To this end, by PCR a portion of the 3'-UTR (untranslated region) of the gene was amplified and sequenced using genomic or cDNA samples as templates which were heterozygous for a T>C variant located in this region (variant ch-v-015 in Table 2A;). As expected, both alleles are represented in the sequence using genomic DNAs as template (Fig. 2B). Both alleles were also equally represented in a sequence of PCR-amplified cDNA from a LE (homozygous wildtype for ch-v-021 and ch-v-26, heterozygous for ch-v-015) liver. In contrast, only the C allele was found in the same portion of *CYP3A5* 3'-UTR cDNA from a HE liver. This indicates an overrepresentation of transcripts derived from the chromosome harbouring the C allele in the total pool of *CYP3A5* transcripts in HE livers.

#### **Example 6: In vitro mutagenesis and expression of recombinant CYP3A5 proteins**

Five polymorphisms in Caucasians have been detected that lead to changes in the protein sequence. Two of them, ch-v-006 and ch-v-017, lead to a truncation of the protein and therefore are unlikely to code for a functional protein. As ch-v-005 has only been found physically linked to ch-v-006, the resulting protein variant is not likely to be functional as well. To determine the effect of the protein variants ch-v-009 and ch-v-001 these variants were analysed in a heterologous bacterial expression system.

A modified CYP3A5 cDNA in the prokaryotic expression vector pKK233-2 (Pharmacia) was used as starting material for in vitro mutagenesis (Fig. 5). Variants ch-v-009 and ch-v-001, respectively, were introduced into the plasmid by in vitro mutagenesis using the QuikChange mutagenesis kit (Stratagene). The successful introduction of the mutations and the absence of other, undesired mutations was confirmed by sequencing. The original plasmid as well as the two mutagenised plasmids were used to transform *E. coli* TOPP3 cells, a strain in which optimal expression of CYP3A proteins has previously been obtained. A total of 8 separate colonies of each mutant plasmid were chosen for expression studies. The bacteria were grown and induced as described in Eiselt *et al.* (Eiselt, Pharmacogenetics 11 (2001), 447-58.). Expression was analysed 48 h and 72 h after induction with IPTG/δ-ALA. Cells were harvested as described in Domanski *et al.*

(Domanski, Arch Biochem Biophys 350 (1998), 223-32). The final P450 content was measured by reduced carbon monoxide (CO) difference spectra (Omura, J. Biol. Chem. 239 (1964), 2370-2378).

Whereas 30 to 50 nmol solubilised CYP3A5 could be recovered per litre culture of the non-mutagenised CYP3A5, expression in the two CYP3A5 variants S100Y and T398N was determined to be lower than 3 nmol P450 protein per litre culture. In many instances, the "P450" peak was shifted to 454 - 458 nm rather than the typical 448 - 450 nm peak expected. The low level of expression in mutagenised colonies made any attempts at protein purification futile. Previous experiments suggest that expression levels as low as those demonstrated by these CYP3A5 variants can not be significantly improved by utilising other bacterial strains or adjusting growth temperature. Therefore, the results strongly suggest that the CYP3A5 protein variants comprising the S100Y or the T398N substitutions are unstable in an *E. coli* expression system and that the variants comprising ch-v-009 or ch-v-001 may not code for functional proteins. The result of negative expression for variant ch-v-001 (T398N) is in agreement with the study in which this polymorphism was initially detected in two of five CYP3A5 deficient individuals (Jounaidi, Biochem Biophys Res Commun 221 (1996), 466-70).

#### **Example 7: Prediction of expected drug metabolism by CYP3A5 genotypes**

The CYP3A5 protein degrades many drugs by oxidation so that they are not therapeutically active anymore. Therefore, drugs that are CYP3A5 substrates might not reach therapeutically active plasma concentrations for an adequate time span in patients with enhanced CYP3A5 activity. In these patients these drugs have to be dosed higher. On the other side, in patients with reduced CYP3A5 activity, these drugs have to be administered at lower dosage in order to avoid toxic drug levels. Table 4 gives an assignment for CYP3A5 genotypes and recommended dosages.

In cases in which CYP3A5 metabolism leads to the formation of pharmacologically active substance, enhanced enzyme activity has to be counteracted by reduced dosage whereas reduced CYP3A5 activity has to be met by increased dosage.

**Table 1A:** Primers used to screen for polymorphisms within the *CYP3A5* upstream regions and the *CYP3A5* gene.

Ref.	Primer			Position (nt)	(bp)
	ID	Name	Sequence (5'-3')		
chzk	001	694	ACAGGCACAGAAACCCACAAG	145448-145468 <sup>1</sup>	630
	002	711	ATCGCCA CTTGCCTTCTTC	146077-146059 <sup>1</sup>	
chzl	003	794	CCCTGCTTCGGCTTGTGCA	159915-159933 <sup>1</sup>	575
	004	750	CACAGCCTGCTTATTGTGATGA	160489-160466 <sup>1</sup>	
chzj	005	751	GATCCTTGGTAGGACAAGCCT	160351-160371 <sup>1</sup>	844
	006	754	CAAGCACTGATTGGTCACCCCT	161194-161171 <sup>1</sup>	
chzb	007	819	GGGATGGGACCGTAAGTGAAC	160951-160972 <sup>1</sup>	618
	008	820	TAATCACATTGGAGTTCTGACAAATG	161568-161543 <sup>1</sup>	
chzi	009	736	AAAAACCTCTACAAAAGTATCATCGATA	161419-161448 <sup>1</sup>	910
	010	737	CCTACTAGGTCTCTGACTTGGAACCAT	162328-162302 <sup>1</sup>	
chzh	011	784	GCCGAGACGCACCATTACACT	161876-161896 <sup>1</sup>	637
	012	785	CACCCATCCCTCCCACTCAT	162512-162492 <sup>1</sup>	
chzg	013	740	TGATGGTTCCAAGTCAGAGACCTAGTAG	162300-162327 <sup>1</sup>	997
	014	741	AATTGTAGACATTTCTCTTAAGTTAATTCCCAG	163296-163262 <sup>1</sup>	
chzf	015	786	TCTGCATGCCAACAGTGAACAATCT	163182-163206 <sup>1</sup>	824
	016	789	GGCACGCACCAGCATGTCC	164005-163987 <sup>1</sup>	
chze	017	790	CTGGCTGAGTGCCGTGGCT	163845-163863 <sup>1</sup>	591
	018	791	TGAGCGCTTCAATGTATTCTGGCTAT	164435-164411 <sup>1</sup>	
chza	019	824	AAATATTTCAAAGTCACACTCTGACAAACAG	164376-164406 <sup>1</sup>	617
	020	822	TAACAGGATCTCATGCTTTTCATGGCT	164992-164964 <sup>1</sup>	
chzd	021	747	CACTCCAATATTACAATAGCCACTATTCA	164843-164872 <sup>1</sup>	926
	022	748	ACTCCTACGTATCCCTCCAAGGCC	165768-165745 <sup>1</sup>	
chzc	023	728	GCTAAGGAAACAGGCATAGAAACTTAC	165586-165613 <sup>1</sup>	557
	024	729	GGAGCTTCCCTGCCCTGC	166142-166125 <sup>1</sup>	
chzy	025	323	TCCTTCTCCAGCACATAAATC	166076-166096 <sup>1</sup>	424
	026	325	AAATTAGAAGGTGGATGGGAG	166499-166479 <sup>1</sup>	
chzx	027	335	GAGTAACTCACCAGCCCTCTG	169838-169858 <sup>1</sup>	264
	028	336	AAACCTCAGAACTCCCTCCCA	170101-170081 <sup>1</sup>	
chzw	029	338	GACATCTCTGAATAGCTTCCCTC	171392-171414 <sup>1</sup>	402
	030	341	GCACATAGTTATAACGGCAA	171793-171773 <sup>1</sup>	
chzv	031	346	AGAACCTAAGGTTGCTGTGTC	173303-173325 <sup>1</sup>	394
	032	348	TGCAAGATGTTACCACTGGGC	173696-173676 <sup>1</sup>	
chzu	033	354	CGCCCCACATAACACTCAGAA	31376-31395 <sup>2</sup>	426
	034	357	AGACCATTTTTAGGAAGCTCG	31801-31781 <sup>2</sup>	
chzt	035	379	CAAGGGTAGTCCACTGAGTTC	31760-31781 <sup>2</sup>	403
	036	381	CTCTTGGAGTTGCAGCG	32162-32145 <sup>2</sup>	
chzs	037	362	AGGTGAGTCTAACTCAGCTTG	33081-33101 <sup>2</sup>	578
	038	365	GACAGCTAAAGTGTGAGGG	33658-33638 <sup>2</sup>	
chzr	039	371	AATGGGTTCCAGTTGAGAACATC	34411-34431 <sup>2</sup>	470
	040	373	ATTGTTGTGCCCTGATTTCAG	34880-34860 <sup>2</sup>	
chzq	041	387	AGAAGCCATAGGGAGGTG	35627-35645 <sup>2</sup>	423
	042	389	GACTGTCCTCCAAGCATTCT	36049-36030 <sup>2</sup>	
chzp	043	394	GATGCCATGATGAGGAGTGTG	37724-37744 <sup>2</sup>	626
	044	397	ACCAGGGCCAGCAATATTG	38349-38331 <sup>2</sup>	
chzm	045	403	AAATACTTCACGAATACTATGATCA	45711-45735 <sup>2</sup>	595
	046	405	CAGGGACATAATTGATTATCTTG	46305-46282 <sup>2</sup>	
chzo	047	411	TACTGGTTGGGAGGTGGAG	48290-48308 <sup>2</sup>	456
	048	412	CATGATGTTCTTAATGCTACAGG	48745-48723 <sup>2</sup>	
chzn	049	419	GAAGAGTTCAAGATACTGGTTA	50088-50112 <sup>2</sup>	416
	050	420	TGCACAAACACTCTACACAGACTC	50503-50481 <sup>2</sup>	

Ref.: Reference sequence. The positions of primers refer to GenBank sequences with accession numbers (1) AF280107.1 and (2) AC005020.2.

**Table 1B:** Primers (Seq IDs 202 and 203) and probes (Seq IDs 204 and 205) used determine the nucleotide status at the polymorphic site ch-v-048 (g.6986G>A) by TaqMan assay.

Ref.	Primer			Position (nt)	(bp)
	ID	Name	Sequence (5'-3')		
chu	202	TQPi_ch-v-048_F	GCTCTACTGTCATTTCTAACATAAT CTCTTTA	173152-173184	99
	204	TQPo_ch-v-048_A1_G_VIC	VIC-TGTCTTCAGTATCTCTT-MGB-DQ	173196-173213	
	205	TQPo_ch-v-048_A12_A_FAM	FAM-TGTCTTCAAATATCTCTTC-MGB-DQ	173196-173214	
	203	TQPi_ch-v-048_R	GCTTCATATGATGAACGGTAATGTGG T	173250-173224	

Ref.: Reference sequence. The positions of the primers refer to the GenBank sequence with accession number AF280107.1. Probes are labelled with a fluorescent dye at the 5' end and labelled with a dark quencher (DQ) and using a minor groove binder (MGB).

**Table 2A: CYP3A5 polymorphisms detected in samples of Caucasian origin.**

Variant ID	Reference sequence	Variant position on		Sequence context		Genetic element	Predicted effect	Caucasian	
		Reference sequence	gDNA	Seq ID	Reference seq. Variant seq.			N	Variant allele frequency (%)
ch-v-020	chzk	254T>G	g.-20619T>G	051	TGGGCCTTGCAA	5' of PS2		211	6.6
ch-v-031	chzk	318G>A	g.-20555G>A	052	.....G.....				
ch-v-032	chzk	544G>A	g.-20329G>A	053	GCAATGGGTAAA	5' of PS2		189	0.3
ch-v-033	chzk	550G>A	g.-20323G>A	054	.....A.....	5' of PS2		186	0.3
ch-v-021	chzl	582A>G	g.-20291A>G	055	GGGGTGTGTGTC	5' of PS2			
ch-v-026	chzl	229A>G	g.-6177A>G	056	.....A.....	5' of PS2		186	0.3
ch-v-027	chzi	566G>A	g.-4336G>A	057	TGTGCCGATTCT	5' of PS2			
ch-v-028	chzh	601G>A	g.-3844G>A	058	.....A.....	5' of PS2		215	6.7
ch-v-029	chzg	464T>C	g.-3557T>C	059	CCCCCACCTCC	5' of PS2		208	7.0
ch-v-034	chza	328T>C	g.-1617T>C	060	CCTCACACTGGG	5' of			
ch-v-030	chzd	683T>A	g.-795T>A	061	ATCCCATGTATA	Exon 1		19	2.7
ch-v-002	chzy	159G>A	g.-866G>A	062	.....G.....	Exon 1		20	10.0
ch-v-003	chzy	171C>T	g.-74C>T	063	GAAGACGCACA	5' of			
ch-v-004	chzy	418-420delGAG	g.1174-176delGAG	064	.....A.....	Exon 1			
ch-v-005	chzx	187C>T	g.3705C>T	065	TCTTATTGCTAT	5' of		93	2.2
ch-v-006	chzx	191-192insG	g.3709-3710insG	066	.....A.....	Exon 1		20	5.0
ch-v-007	chzw	143C>T	g.5215C>T	067	CACTCTAACCC	5' of			
				068	.....C.....	Exon 1			
				069	.....T.....	Exon 1			
				070	.....A.....	Exon 1			
				071	.....A.....	Exon 1			
				072	.....A.....	Exon 1			
				073	GGCGAGGGAAAGC	Exon 1			
				074	.....A.....	(5' UTR)			
				075	CCAGGGAAACA	Exon 1			
				076	.....T.....	(5' UTR)			
				077	TCAAGGAGAAAG	Intron 1			
				078	.....T.....				
				079	GTAACACATGGA	Exon 2	H30Y	104	1.4
				080	.....G.....	Exon 2	(K34)¹	104	1.4
				081	CATGG-ACTTT				
				082	.....G.....				
				083	GATAGCAGGCC				
				084	.....T.....				

ch-v-048	chyu	206G>A	g.6986G>A	147	.....A.....	Inton 3	splice defect	217	4.6
ch-v-008	chzv	199C>A	g.7182C>A	085	AGAATCGGGCT	Inton 3		107	1.9
ch-v-009	chzv	320C>A	g.7303C>A	086	.....A.....	Inton 3			
ch-v-018	chzv	441-444insCTAAAAAAAT	g.7424-7427insCTAAAAAAAT	087	TTATTCTGTCT	Exon 4	S100Y	107	0.5
ch-v-016	chzt	145T>G	g.13077T>G	088	.....A.....	Inton 4			
ch-v-019	chzt	241T>C	g.13173T>C	089	C---AG----G	Inton 4		107	0.5
ch-v-010	chzq	132-133insGTC	g.16931-16932insGTC	090	CCTAAAAAAATG				
ch-v-011	chzq	364G>T	g.17163G>T	091	TCTTTTTATCTT	Inton 5		105	0.5
ch-v-012	chzp	269G>A	g.19165G>A	092	.....G.....	Inton 5			
ch-v-013	chzm	167A>G	g.27050A>G	093	GAGTCCTGCACA	Inton 5		105	0.5
ch-v-001	chzm	406C>A	g.27289C>A	094	.....C.....	Inton 5			
ch-v-014	chzm	643C>T	g.27526C>T	095	AGTC---AAGA	Inton 8		95	0.5
ch-v-015	chzn	351T>C	g.31611T>C	096	.....GTC.....	Inton 8			
				097	AGGAAGTATTC	Inton 9		95	0.5
				098	.....T.....	Inton 9			
				099	AGAGAGCTCA	Inton 9		106	0.5
				100	.....A.....	Inton 9			
				101	CTTCAATAAGTA	Inton 10		80	11.9
				102	.....G.....	Inton 10			
				103	TCCAACCTATG	Exon 11	T398N	80	1.9
				104	.....A.....	Exon 11			
				105	CGAAAACTACAT	Inton 11		80	3.8
				106	.....T.....	Inton 11			
				107	AAGGATTCTA	Exon 13		197	5.6
				108	.....C.....	(3' UTR)			

All variants were detected in the heterozygous state except for 1 homozygous individual for variant ch-v-013. Variant results in a frame shift which ultimately leads to a premature termination.

**Table 2B:** *CYP3A5* polymorphisms detected in samples of African-American origin.

Variant ID	Reference sequence	Variant position on		Sequence context		Genetic element	Predicted effect	N	Variant allele frequency (%)
		Reference sequence	gDNA	Seq ID	reference seq. variant seq.				
ch-v-037	chzk	230T>C	g.-20643T>C	148	TTTAATAGAACG.....C.....	5' of PS2		42	3.6
ch-v-020	chzk	254T>G	g.-20619T>G	149	.....C.....	5' of PS2		41	69.5
ch-v-038	chzk	506C>T	g.-20367C>T	051	TGGGCTTGCAA.....G.....	5' of PS2		44	1.1
ch-v-039	chzk	514T>C	g.-20359T>C	052	.....G.....	5' of PS2		44	1.1
ch-v-021	chzk	582A>G	g.-20291A>G	150	ATTCCCCATAG.....T.....	5' of PS2		44	
ch-v-026	chzl	229A>G	g.-6177A>G	151	.....T.....	5' of PS2		44	
ch-v-051	chzh	455T>G	g.-39990T>G	152	TAGAAATATGAA.....C.....	5' of PS2		44	
ch-v-052	chzh	577G>A	g.-38668G>A	153	.....C.....	5' of PS2		45	66.7
ch-v-028	chzh	601G>A	g.-38444G>A	059	GCCCACCCCTCC.....G.....	5' of PS2		45	
ch-v-034	chyz	328T>C	g.-16177T>C	060	.....G.....	5' of PS2		43	65.1
ch-v-002	chzy	159G>A	g.-86G>A	061	CTCACACTGGGG.....C.....	5' of PS2		43	
ch-v-003	chzy	171C>T	g.-74C>T	062	.....G.....	5' of PS2		44	
ch-v-007	chzw	143C>T	g.5215C>T	063	GTAACITATGCC.....G.....	5' of PS2		44	2.3
ch-v-053	chzw	163G>A	g.5235G>A	064	CATCTTACCCCC.....G.....	5' of PS2		45	3.4
ch-v-054	chzw	444T>A	g.5516T>A	065	TTCACGTGGAG.....A.....	5' of PS2		44	17.1
ch-v-048	chyu	206G>A	g.6986G>A	066	TGTGTGTGGGA.....A.....	5' of PS2		44	
				067	.....A.....	5' of PS2		45	
				068	.....A.....	5' of PS2		45	
				069	CATCTTACCCCC.....C.....	5' of PS2		45	
				070	.....C.....	5' of PS2		45	
				071	GGCAAGGGAAAGC.....A.....	5' of PS2		45	
				072	.....A.....	5' of PS2		45	
				073	CCAGGCATAACA.....T.....	5' of PS2		45	
				074	.....A.....	5' of PS2		45	
				075	.....T.....	5' of PS2		45	
				076	.....A.....	5' of PS2		45	
				077	.....T.....	5' of PS2		44	1.1
				078	.....A.....	5' of PS2		44	
				079	.....T.....	5' of PS2		44	
				080	.....A.....	5' of PS2		44	
				081	.....T.....	5' of PS2		44	
				082	.....A.....	5' of PS2		44	
				083	.....T.....	5' of PS2		44	
				084	.....A.....	5' of PS2		44	
				085	.....T.....	5' of PS2		44	
				086	.....A.....	5' of PS2		44	
				087	.....T.....	5' of PS2		44	
				088	.....A.....	5' of PS2		44	
				089	.....T.....	5' of PS2		44	
				090	.....A.....	5' of PS2		44	
				091	.....T.....	5' of PS2		44	
				092	.....A.....	5' of PS2		44	
				093	.....T.....	5' of PS2		44	
				094	.....A.....	5' of PS2		44	
				095	.....T.....	5' of PS2		44	
				096	.....A.....	5' of PS2		44	
				097	.....T.....	5' of PS2		44	
				098	.....A.....	5' of PS2		44	
				099	.....T.....	5' of PS2		44	
				100	.....A.....	5' of PS2		44	
				101	.....T.....	5' of PS2		44	
				102	.....A.....	5' of PS2		44	
				103	.....T.....	5' of PS2		44	
				104	.....A.....	5' of PS2		44	
				105	.....T.....	5' of PS2		44	
				106	.....A.....	5' of PS2		44	
				107	.....T.....	5' of PS2		44	
				108	.....A.....	5' of PS2		44	
				109	.....T.....	5' of PS2		44	
				110	.....A.....	5' of PS2		44	
				111	.....T.....	5' of PS2		44	
				112	.....A.....	5' of PS2		44	
				113	.....T.....	5' of PS2		44	
				114	.....A.....	5' of PS2		44	
				115	.....T.....	5' of PS2		44	
				116	.....A.....	5' of PS2		44	
				117	.....T.....	5' of PS2		44	
				118	.....A.....	5' of PS2		44	
				119	.....T.....	5' of PS2		44	
				120	.....A.....	5' of PS2		44	
				121	.....T.....	5' of PS2		44	
				122	.....A.....	5' of PS2		44	
				123	.....T.....	5' of PS2		44	
				124	.....A.....	5' of PS2		44	
				125	.....T.....	5' of PS2		44	
				126	.....A.....	5' of PS2		44	
				127	.....T.....	5' of PS2		44	
				128	.....A.....	5' of PS2		44	
				129	.....T.....	5' of PS2		44	
				130	.....A.....	5' of PS2		44	
				131	.....T.....	5' of PS2		44	
				132	.....A.....	5' of PS2		44	
				133	.....T.....	5' of PS2		44	
				134	.....A.....	5' of PS2		44	
				135	.....T.....	5' of PS2		44	
				136	.....A.....	5' of PS2		44	
				137	.....T.....	5' of PS2		44	
				138	.....A.....	5' of PS2		44	
				139	.....T.....	5' of PS2		44	
				140	.....A.....	5' of PS2		44	
				141	.....T.....	5' of PS2		44	
				142	.....A.....	5' of PS2		44	
				143	.....T.....	5' of PS2		44	
				144	.....A.....	5' of PS2		44	
				145	.....T.....	5' of PS2		44	
				146	.....A.....	5' of PS2		44	
				147	.....T.....	5' of PS2		44	
				148	.....A.....	5' of PS2		44	
				149	.....T.....	5' of PS2		44	
				150	.....A.....	5' of PS2		44	
				151	.....T.....	5' of PS2		44	
				152	.....A.....	5' of PS2		44	
				153	.....T.....	5' of PS2		44	
				154	.....A.....	5' of PS2		44	
				155	.....T.....	5' of PS2		44	
				156	.....A.....	5' of PS2		44	
				157	.....T.....	5' of PS2		44	
				158	.....A.....	5' of PS2		44	
				159	.....T.....	5' of PS2		44	
				160	.....A.....	5' of PS2		44	
				161	.....T.....	5' of PS2		44	
				162	.....A.....	5' of PS2		44	
				163	.....T.....	5' of PS2		44	
				164	.....A.....	5' of PS2		44	
				165	.....T.....	5' of PS2		44	
				166	.....A.....	5' of PS2		44	
				167	.....T.....	5' of PS2		44	
				168	.....A.....	5' of PS2		44	
				169	.....T.....	5' of PS2		44	
				170	.....A.....	5' of PS2		44	
				171	.....T.....	5' of PS2		44	
				172	.....A.....	5' of PS2		44	
				173	.....T.....	5' of PS2		44	
				174	.....A.....	5' of PS2		44	
				175	.....T.....	5' of PS2		44	
				176	.....A.....	5' of PS2		44	
				177	.....T.....	5' of PS2		44	
				178	.....A.....	5' of PS2		44	
				179	.....T.....	5' of PS2		44	
				180	.....A.....	5' of PS2		44	
				181	.....T.....	5' of PS2		44	
				182	.....A.....	5' of PS2		44	
				183	.....T.....	5' of PS2		44	
				184	.....A.....	5' of PS2		44	
				185	.....T.....	5' of PS2		44	
				186	.....A.....	5' of PS2		44	
				187	.....T.....	5' of PS2		44	
				188	.....A.....	5' of PS2		44	
				189	.....T.....	5' of PS2		44	
				190	.....A.....	5' of PS2		44	
				191	.....T.....	5' of PS2		44	
				192	.....A.....	5' of PS2		44	
				193	.....T.....	5' of PS2		44	
				194	.....A.....	5' of PS2		44	
				195	.....T.....	5' of PS2		44	
				196	.....A.....	5' of PS2		44	
				197	.....T.....	5' of PS2		44	
				198	.....A.....	5' of PS2		44	
				199	.....T.....	5' of PS2		44	
				200	.....A.....	5' of PS2		44	
				201	.....T.....	5' of PS2		44	
				202	.....A.....	5' of PS2		44	
				203	.....T.....	5' of PS2		44	
				204	.....A.....	5' of PS2		44	
				205	.....T.....	5' of PS2		44	
				206	.....A.....	5' of PS2		44	
				207	.....T.....	5' of PS2		44	
				208	.....A.....	5' of PS2		44	
				209	.....T.....	5' of PS2		44	
				210	.....A.....	5' of PS2		44	
				211	.....T.....	5' of PS2		44	
				212	.....A.....	5' of PS2		44	
				213	.....T.....	5' of PS2		44	
				214	.....A.....	5' of PS2		44	
				215	.....T.....	5' of PS2		44	
				216	.....A.....	5' of PS2		44	
				217	.....T.....	5' of PS2		44	
				218	.....A.....	5' of PS2		44	
				219	.....T.....	5' of PS2		44	
				220	.....A.....	5' of PS2		44	
				221	.....T.....	5' of PS2		44	
				222	.....A.....	5' of PS2		44	
				223	.....T.....	5' of PS2		44	
				224	.....A.....	5' of PS2		44	
				225	.....T.....	5' of PS2		44	
				226	.....A.....	5' of PS2		44	
				227	.....T.....	5' of PS2		44	
				228	.....A.....	5' of PS2		44	
				229	.....T.....	5' of PS2		44	
				230	.....A.....	5' of PS2		44	
				231	.....T.....	5' of PS2		44	
				232	.....A.....	5' of PS2		44	
				233	.....T.....	5' of PS2		44	
				234	.....A.....	5' of PS2			

ch-v-025	chzv	224C>T	g.7207C>T	109 110 .....T.....	AGCTCCGTTGT CATCAATTGCC Exon 6	Intron 3	43	7.0
ch-v-043	chzt	294T>C	g.13226T>C	162 163 .....C.....	CAGTCGCAC TG Intron 6	I149T	45	1.1
ch-v-055	chzt	444G>A	g.13376G>A	164 165 .....A.....	ACTAAGAACGGT .....A.....	Intron 6	45	1.1
ch-v-050	chzs	437G>A	g.14690G>A	166 167 .....A.....	ACTAAGAACGGT .....A.....	Exon 7	splice defect	13.3
ch-v-056	chzs	467A>G	g.14720A>G	168 169 .....G.....	GATCCATTATT .....G.....	Exon 7	P218P	45
ch-v-057	chzs	583C>T	g.14836C>T	170 171 .....T.....	CAAATCCATGT .....T.....	Intron 7	45	1.1
ch-v-058	chzs	650A>G	g.14903A>G	172 173 .....G.....	TGTCAATCTAG .....G.....	Intron 7	45	6.7
ch-v-059	chzr	205T>C	g.15788T>C	174 175 .....C.....	TGGTTTGTTTT .....C.....	Intron 7	44	3.4
ch-v-060	chzr	496A>C	g.16079A>C	176 177 .....C.....	AAATAAAAGAAC .....C.....	Intron 8	44	1.1
ch-v-011	chzq	364G>T	g.17163G>T	097 098 .....T.....	AGGAAGTATTC .....T.....	Intron 9	43	7.0
ch-v-062	chzp	173G>A	g.19069G>A	178 179 .....A.....	TGGCGGTCACTC .....A.....	Intron 9	45	1.1
ch-v-063	chzp	312C>T	g.19208C>T	180 181 .....T.....	TTGACCTGATT .....T.....	Intron 9	45	2.2
ch-v-013	chzm	167A>G	g.27050A>G	101 102 .....G.....	CTTCAAATAGTA .....G.....	Intron 10	36	1.4
ch-v-017	chzm	248-249insT	g.27131-27132insT	111 112 .....T.....	CACCTT-ACCTA .....T.....	Exon 11 (D348.) <sup>1</sup>	45	10.0
ch-v-014	chzm	643C>T	g.27526C>T	105 106 .....T.....	CGAAACTACAT .....T.....	Intron 11	42	11.9
ch-v-044	chzn	239T>C	g.31499T>C	182 183 .....C.....	TATTGTAGATC .....C.....	Intron 12	45	5.6
ch-v-015	chzn	351T>C	g.31611T>C	107 108 .....C.....	AAGGATTCTA .....C.....	Exon 13 (3' UTR)	44	68.2

<sup>1</sup> variant results in a frame shift which ultimately leads to premature termination.

**Table 2C: CYP3A5 polymorphisms detected in samples of Chinese origin.**

Variant ID	Reference sequence	Variant position on		Sequence context		Genetic element	Predicted effect	N	Variant allele frequency (%)
		Reference sequence	gDNA	Seq ID	reference seq. variant seq.				
ch-v-020	chzk	254T>G	g.-20619T>G	051	TGGGCTTCGAA	5' of PS2		42	23.8
ch-v-021	chzk	582A>G	g.-20291A>G	052	.....G.....	5' of PS2		45	26.7
ch-v-026	chzl	229A>G	g.-6177A>G	060	.....G.....	5' of PS2			
ch-v-034	chyz	328T>C	g.-16177T>C	061	CCTCACACTGGG	5' of Exon 1		47	26.6
ch-v-066	chzy	380T>C	g.1386T>C	062	.....G.....	5' of Exon 1			
ch-v-067	chzy	474G>A	g.230G>A	069	CATCTTACCCC	5' of Exon 1		47	21.3
ch-v-048	chyu	206G>A	g.6986G>A	184	CCPTTTTCCCTT	Intron 1		45	1.1
ch-v-018	chzv	441-444insCTAAAAAAAT	9.7424-7427insCTAAAAAAAT	089	C---AG-----G	Intron 4		44	1.14
ch-v-068	chzu	359G>A	g.12907G>A	090	CCTAAAAAAATG	Intron 4		47	2.1
ch-v-047	chzu	404T>C	g.12952T>C	188	AATAACGGTCAT	Exon 5	R130Q	45	3.3
ch-v-069	chzu	480G>A	g.13028G>A	189	.....A.....	Intron 5	splice defect	44	1.1
ch-v-011	chzq	364G>T	g.17163G>T	190	GAGGGTATGAA	Intron 5		44	1.1
ch-v-014	chzm	643C>T	g.27526C>T	191	.....C.....	Intron 5			
ch-v-015	chzn	351T>C	g.31611T>C	192	AGTCGGTTTC	Intron 5			
				193	.....A.....	Intron 5			
				097	AGGAAGTATTTC	Intron 9		40	21.3
				098	.....T.....	Intron 9			
				105	CGAAACTACAT	Intron 11		47	5.3
				106	.....T.....	Intron 11			
				107	ANGGATTTCATA	Exon 13 (3' UTR)		47	26.6
				108	.....C.....	Exon 13 (3' UTR)			

**Table 2D: CYP3A5 polymorphisms detected in samples of Japanese origin.**

Variant ID	Reference sequence	Variant position on		Sequence context		Genetic element	Predicted effect	N	Variant allele frequency (%)
		Reference sequence	gDNA	Seq ID	reference seq. variant seq.				
ch-v-020	chzk	254T>G	g.-20619T>G	051	TGGGCCTTGCAA	5' of PS2		42	29.8
ch-v-021	chzk	582A>G	g.-20291A>G	052	.....G.....				
ch-v-026	chzl	229A>G	g.-6177A>G	059	GCCCCACCTCC	5' of PS2		49	28.6
ch-v-028	chzh	601G>A	g.-3844G>A	060	.....G.....				
ch-v-034	chyz	328T>C	g.-1617T>C	061	C'TCACACTGGG	5' of		46	28.3
ch-v-007	chzw	143C>T	g.5215C>T	062	.....G.....	Exon 1			
ch-v-048	chyu	206G>A	g.6986G>A	065	TGTGTGTGGGA	5' of		50	2.0
ch-v-047	chzu	404T>C	g.12952T>C	066	.....A.....	Exon 1			
ch-v-061	chzq	194C>G	g.16993C>G	069	CATCTTACCCC	5' of		50	26.0
ch-v-011	chzq	364G>T	g.17163G>T	070	.....C.....	Exon 1			
ch-v-017	chzm	248-249insT	g.27131-27132insT	083	GATAAGCAGGCC	Intron 2		48	3.1
ch-v-014	chzm	643C>T	g.27526C>T	084	.....T.....				
ch-v-045	chzn	291T>C	g.31551T>C	098	TTTCAGATATCT	Intron 3		50	29.0
ch-v-015	chzn	351T>C	g.31611T>C	109	GAGGGATATGAA	Intron 5	splice defect	50	1.0
				191	.....C.....				
				194	TCTGCCAAAGA	Intron 8		49	1.0
				195	.....G.....				
				097	AGGAAGTATTC	Intron 9			
				098	.....T.....			49	
				111	CACCT-ACCTA	Exon 11	(D348.) <sup>1</sup>	48	1.0
				112	.....T.....				
				105	CGAAACTACAT	Intron 11		49	3.1
				106	.....T.....				
				196	ACCCATTGTC	Exon 13	I488T	50	3.0
				197	.....C.....				
				107	AAGGATTCTTA	Exon 13 (3' UTR)		50	31.0
				108	.....C.....				

**Table 2E:** CYP3A5 polymorphisms detected in samples of Korean origin.

Variant ID	Reference sequence	Variant position on reference sequence	gDNA	Seq ID	reference seq. variant seq.	Sequence context	Genetic element	Predicted effect	N	Variant allele frequency (%)
ch-v-020	chzk	254T>G	g.-20619T>G	051	TGGGCCTTGCAA	5' of PS2			47	29.8
ch-v-065	chzk	563T>C	g.-20310T>C	052	.....G.....					
ch-v-021	chzk	582A>G	g.-20291A>G	198	GCTACTGGCTG	5' of PS2			47	1.1
ch-v-040	chzl	206C>T	g.-6200C>T	199	.....C.....					
ch-v-026	chzl	229A>G	g.-6177A>G	059	GCCCCACCTCC	5' of PS2			47	29.8
ch-v-034	chyz	328T>C	g.-1617T>C	060	.....G.....					
ch-v-007	chzw	143C>T	g.5215C>T	200	GAAATCACCCG	5' of			43	1.2
ch-v-048	chu	206G>A	g.6986G>A	201	.....T.....	Exon 1				
ch-v-061	chzq	194C>G	g.16993C>G	061	CTCACACTGG	5' of			43	31.4
ch-v-011	chzq	364G>T	g.17163G>T	062	.....G.....	Exon 1				
ch-v-014	chzm	643C>T	g.27526C>T	063	GATAGCAGGCC	5' of			47	27.7
ch-v-015	chzn	351T>C	g.31611T>C	084	.....T.....	Intron 2				
				146	TTTCAGTATCT	Intron 3			47	29.8
				147	.....A.....					
				194	TCTGCCAAAGA	Intron 8			47	2.1
				195	.....G.....					
				097	AGGAAGTATC	Intron 9			47	27.7
				098	.....T.....					
				105	CGAAACTACAT	Intron 11			47	2.1
				106	.....T.....					
				107	AAGGATTCTA	Exon 13			43	36.1
				108	.....C.....	(3' UTR)				

Table 2A-E: Variants are listed according to their localisation along the gene, separately for each ethnic group. Polymorphism nomenclature is based on Antoarakis, Hum Mutat 11 (1998), 1-3) using the joined sequences AF280107.1 and AC005020.2 as genomic reference sequences wherein the A of the ATG at position 166220 in AF280107.1 is +1. Sequence context: local alignment at the polymorphic site with the reference allele sequence given at the top and the variant sequence given below. Dots indicate nucleotide identity at the respective position. N: number of samples analysed.

**Table 3: CYP3A5 genotypes and phenotypes.**

	ch-v-021	ch-v-026	ch-v-015	Phenotype	Livers
Genotype 1	A/A	A/A	T/T	LE	155
Genotype 2	A/G	A/G	T/T	LE	9
Genotype 3	A/A	A/A	T/C	LE	4
Genotype 4	A/G	A/G	T/C	HE	17
Genotype 5	A/A	A/A	T/T	HE	1

All three variants were observed only in the heterozygous state. HE = high expressing livers, LE = low expressing livers. Numbers indicate LE and HE livers with each particular genotype. The increased CYP3A5 expression co-segregates with a distinct genotype.

Table 4: Expected drug metabolism by CYP3A5

Geno-type No.	Allelic combination	Enzyme Activity	Dose Adjustment	
			drug degradation	drug activation
I	high expressor allele/ high expressor allele	190 %	1.90	0.53
II, III	low expressor allele/ high expressor allele	100 %	1.00	1.00
IV - VII	null allele/ high expressor allele	95 %	0.95	1.05
VIII, IX, X	low expressor allele/ low expressor allele	10 %	0.10	10
XI - XVIII	null allele/ low expressor allele	5 %	0.05	20
XIX - XXV	null allele/ null allele	0 %	< 0.05	> 20
Geno-type No.	Genotype (SeqID) at Locus 1 - 2 - 3 - 4 - 5 - 6 - 7 - 8	Enzyme Activity	Dose Adjustment	
			drug degradation	drug activation
I	060-062-079-081-087-111-103-108 / 060-062-079-081-087-111-103-108	190 %	1.90	0.53
II	0xx-06x-079-081-087-111-103-107 / 060-062-079-081-087-111-103-108	100 %	1.00	1.00
III	059-061-079-081-087-111-103-10x / 060-062-079-081-087-111-103-108	100 %	1.00	1.00
IV	0xx-06x-080-082-08x-11x-10x-10x / 060-062-079-081-087-111-103-108	95 %	0.95	1.05
V	0xx-06x-0xx-08x-088-11x-10x-10x / 060-062-079-081-087-111-103-108	95 %	0.95	1.05
VI	0xx-06x-0xx-08x-08x-112-10x-10x / 060-062-079-081-087-111-103-108	95 %	0.95	1.05
VII	0xx-06x-0xx-08x-08x-11x-104-10x / 060-062-079-081-087-111-103-108	95 %	0.95	1.05
VIII	060-062-079-081-087-111-103-107 / 059-061-079-081-087-111-103-108	10 %	0.10	10
IX	0xx-06x-079-081-087-111-103-107 / 0xx-06x-079-081-087-111-103-107	10 %	0.10	10
X	059-061-079-081-087-111-103-10x / 059-061-079-081-087-111-103-10x	10 %	0.10	10
XI	059-061-079-081-087-111-103-10x / 0xx-06x-080-082-08x-11x-10x-10x	5 %	0.05	20
XII	0xx-06x-079-081-087-111-103-107 / 0xx-06x-080-082-08x-11x-10x-10x	5 %	0.05	20
XIII	059-061-079-081-087-111-103-10x / 0xx-06x-0xx-08x-088-11x-10x-10x	5 %	0.05	20
XIV	0xx-06x-079-081-087-111-103-107 / 0xx-06x-0xx-08x-088-11x-10x-10x	5 %	0.05	20
XV	059-061-079-081-087-111-103-10x / 0xx-06x-0xx-08x-08x-112-10x-10x	5 %	0.05	20
XVI	0xx-06x-079-081-087-111-103-107 /	5 %	0.05	20

	0xx-06x-0xx-08x-08x-112-10x-10x			
XVII	059-061-079-081-087-111-103-10x / 0xx-06x-0xx-08x-08x-11x-104-10x	5 %	0.05	20
XVIII	0xx-06x-079-081-087-111-103-107 / 0xx-06x-0xx-08x-08x-11x-104-10x	5 %	0.05	20
XIX	0xx-06x-080-082-08x-11x-10x-10x / 0xx-06x-0xx-08x-088-11x-10x-10x	0 %	< 0.05	> 20
XX	0xx-06x-080-082-08x-11x-10x-10x / 0xx-06x-0xx-08x-08x-112-10x-10x	0 %	< 0.05	> 20
XXI	0xx-06x-080-082-08x-11x-10x-10x / 0xx-06x-0xx-08x-08x-11x-104-10x	0 %	< 0.05	> 20
XXII	0xx-06x-080-082-08x-11x-10x-10x / 0xx-06x-080-082-08x-11x-10x-10x	0 %	< 0.05	> 20
XXIII	0xx-06x-0xx-08x-088-11x-10x-10x / 0xx-06x-0xx-08x-088-11x-10x-10x	0 %	< 0.05	> 20
XXIV	0xx-06x-0xx-08x-08x-112-10x-10x / 0xx-06x-0xx-08x-08x-112-10x-10x	0 %	< 0.05	> 20
XXV	0xx-06x-0xx-08x-08x-11x-104-10x / 0xx-06x-0xx-08x-08x-11x-104-10x	0 %	< 0.05	> 20

No.: running genotype number.

**Genotype:** Possible CYP3A5 genotypes that result from combinations of alleles. At the top of the table concise allele names have been used to indicate the principle. The lower table lists the alleles in greater detail, giving all combinations of variants at 8 loci in the two homologous chromosomes (loci 1 -8 refer to positions corresponding to positions -20291, -6177, 3705, 3709/3710, 7303, 27131/27132, 27289 and 31611, respectively, of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613), respectively. Each variant is defined by a 3-digit Seq ID as listed in Table 2A-E. A wildcard (x) in Seq IDs indicates that the phenotype is independent from the variant at this locus in this chromosome. The possible variants for each locus that can be substituted for x can be extracted from Table 2A-E. For example, 08x at locus 4 stands for Seq IDs 081 or 082, whereas 08x at locus 5 indicates either Seq ID 087 or 088.

**Enzyme Activity:** enzyme activity as calculated from protein concentration whereby the average protein concentration of genotype 059-061-079-081-111-107/060-062-079-081-111-108 was defined as 100 %.

**Dose Adjustment:** dose adjustment factors for drugs that are degraded/activated by CYP3A5 relative to the dosis required for genotype 059-061-079-081-111-107/060-062-079-081-111-108. Factors may need to be weighted according to the activity share of the CYP3A5 enzyme for drugs which are not exclusively metabolised by CYP3A5.

**CLAIMS**

1. A polynucleotide comprising a polynucleotide selected from the group consisting of:
  - (a) a polynucleotide having the nucleic acid sequence of SEQ ID NO: 54, 56, 58, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 106, 108, 110, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 128, 129, 130, 131, 133, 134, 135, 136, 137, 138, 139, 140, 142, 143, 149, 151, 153, 155, 157, 159, 161, 163, 165, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 193, 195, 197, 199, 201, 207, 208, 209, 210, 211, 212, 213, 214, 216, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 231, 232, 233, 235, or 236;
  - (b) a polynucleotide encoding a polypeptide having the amino acid sequence of SEQ ID NO: 127, 132, 141, 215, 229, or 234;
  - (c) a polynucleotide capable of hybridizing to a CYP3A5 gene, wherein said polynucleotide is having a nucleotide exchange, a nucleotide deletion of at least one nucleotide, or at least one additional nucleotide at a position corresponding to position -20643, -20555, -20359, -20367, -20329, -20323, -20310, -6200, -6177, -4336, -3990, -3868, -3844, -3557, -1617, -795, -86, -74, 136, 174 to 176, 230, 3705, 3709/3710, 5215, 5235, 5516, 7182, 7207, 7303, 7424/7427, 12907, 13028, 13077, 13173, 13226, 13376, 14720, 14836, 14903, 15788, 16079, 16931/16932, 16993, 17163, 19069, 19165, 19208, 27050, 27131/27132, 27526, 31499, 31551 or 31611;
  - (d) a polynucleotide capable of hybridizing to a CYP3A5 gene, wherein said polynucleotide is having an A at a position corresponding to position -20555, -20329, -20323, -4336, -3868, -3844, -795, -86, 230, 5235, 5516, 7182, 7303, 12907, 13028, 13376, 19069 or 19165 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614), a T at a position corresponding to position -20367, -6200, -74, 3705, 5215, 7207, 14836, 17163, 19208 or 27526 of

the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614), a G at a position corresponding to position -6177, -3990, 13077, 14720, 14903, 16993 or 27050 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614), a C at a position corresponding to position -20643, -20310, -3557, -1617, 136, 13173, 13226, 15788, 16079, 31499, 31551 or 31611 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614), nucleotide deletions at positions corresponding to positions 174 to 176 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613), an additional nucleotide at a position corresponding to position 3709/3710 or 27131/27132 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614), three additional nucleotides at a position corresponding to position 16931/16932 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614), or a deletion of two nucleotides and nine additional nucleotides inserted at a position corresponding to position 7424 to 7427 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613);

(e) a polynucleotide encoding a CYP3A5 polypeptide or fragment thereof, wherein said polypeptide comprises an amino acid substitution at a position corresponding to position 30, 100, 130, 149 or 488 of the

CYP3A5 polypeptide (Accession No: NP\_000768.1), or at least one amino acid exchange or a stop codon at a position corresponding to position 30 to 34 or 346 to 348 of the CYP3A5 polypeptide (Accession No: NP\_000768.1); and

- (f) a polynucleotide encoding a CYP3A5 polypeptide or fragment thereof, wherein said polypeptide comprises amino acid substitutions of HGLFK to YGTF. (with the period meaning termination) at a position corresponding to position 30 to 34 of the CYP3A5 polypeptide (Accession No: NP\_000768.1), an amino acid substitution of S to Y at a position corresponding to position 100 of the CYP3A5 polypeptide (Accession No: NP\_000768.1), an amino acid substitution of R to Q at a position corresponding to position 130 of the CYP3A5 polypeptide (Accession No: NP\_000768.1), an amino acid substitution of I to T at a position corresponding to position 149 of the CYP3A5 polypeptide (Accession No: NP\_000768.1), an amino acid substitutions of TYD to YL. (with the period meaning termination) at position corresponding to position 346 to 348 of the CYP3A5 polypeptide (Accession No: NP\_000768.1), or an amino acid substitution of I to T at a position corresponding to position 488 of the CYP3A5 polypeptide (Accession No: NP\_000768.1).

2. A polynucleotide of claim 1, wherein said polynucleotide is associated with cancer or diseases including cardiovascular diseases, diabetes and AIDS.
3. A polynucleotide of any one of claims 1 to 2 which is DNA or RNA.
4. A gene comprising the polynucleotide of any one of claims 1 to 2.
5. The gene of claim 4 wherein a nucleotide deletion, addition and/or substitution results in altered expression of the variant gene compared to the corresponding wild type gene.
6. A vector comprising a polynucleotide of any one of claims 1 to 3 or the gene of claim 4 or 5.

7. The vector of claim 6, wherein the polynucleotide is operatively linked to expression control sequences allowing expression in prokaryotic or eukaryotic cells or isolated fractions thereof.
8. A host cell genetically engineered with the polynucleotide of any one of claims 1 to 3, the gene of claim 4 or 5 or the vector of claim 6 or 7.
9. A method for producing a molecular variant CYP3A5 polypeptide or fragment thereof comprising
  - (a) culturing the host cell of claim 8; and
  - (b) recovering said protein or fragment from the culture.
10. A method for producing cells capable of expressing a molecular variant CYP3A5 polypeptide comprising genetically engineering cells with the polynucleotide of any one of claims 1 to 3, the gene of claim 4 or 5 or the vector of claim 6 or 7.
11. A polypeptide or fragment thereof encoded by the polynucleotide of any one of claims 1 to 3, the gene of claim 4 or 5 or obtainable by the method of claim 9 or from cells produced by the method of claim 10.
12. An antibody which binds specifically to the polypeptide of claim 11.
13. The antibody of claim 12 which specifically recognizes an epitope containing one or more amino acid substitution(s) resulting from a nucleotide exchange as defined in claim 1 or 5.
14. The antibody of claim 12 or 13 which is monoclonal or polyclonal.
15. A transgenic non-human animal comprising at least one polynucleotide of any one of claims 1 to 4, the gene of claim 5 or 6 or the vector of claim 7 or 8.

16. The transgenic non-human animal of claim 15 which is a mouse, a rat or a zebrafish.
17. A solid support comprising one or a plurality of the polynucleotide of any one of claims 1 to 3, the gene of claim 4 or 5, the vector of claim 6 or 7, the polypeptide of claim 11, the antibody of claim 12 or 13 or the host cell of claim 8 in immobilized form.
18. The solid support of claim 17, wherein said solid support is a membrane, a glass-, polypropylene- or silicon-chip, are oligonucleotide conjugated beads or bead array, which is assembled on an optical filter substrate.
19. An in vitro method for identifying a polymorphism said method comprising the steps of:
  - (a) isolating a polynucleotide of any one claims 1 to 3 or the gene of claim 4 or 5 from a plurality of subgroups of individuals, wherein one subgroup has no prevalence for a CYP3A5 associated disease and at least one or more further subgroup(s) do have prevalence for a CYP3A5 associated disease; and
  - (b) identifying a polymorphism by comparing the nucleic acid sequence of said polynucleotide or said gene of said one subgroup having no prevalence for a CYP3A5 associated disease with said at least one or more further subgroup(s) having a prevalence CYP3A5 associated disease.
20. A method for identifying and obtaining a pro-drug or a drug capable of modulating the activity of a molecular variant of a CYP3A5 polypeptide comprising the steps of:
  - (a) contacting the polypeptide of claim 11, the solid support of claim 17 or 18, a cell expressing a molecular variant gene comprising a polynucleotide of any one of claims 1 to 3, the gene of claim 4 or 5 or the vector of claim 6 or 7 in the presence of components capable of providing a detectable signal in response to drug activity with a compound to be screened for pro-drug or drug activity; and

- (c) detecting the presence or absence of a signal or increase or decrease of a signal generated from the pro-drug or the drug activity, wherein the absence, presence, increase or decrease of the signal is indicative for a putative pro-drug or drug.
21. A method for identifying and obtaining an inhibitor of the activity of a molecular variant of a CYP3A5 polypeptide comprising the steps of:
- (d) contacting the protein of claim 11, the solid support of claim 17 or 18 or a cell expressing a molecular variant gene comprising a polynucleotide of any one of claims 1 to 3 or the gene of claim 4 or 5 or the vector of claim 6 or 7 in the presence of components capable of providing a detectable signal in response to drug activity with a compound to be screened for inhibiting activity; and
  - (e) detecting the presence or absence of a signal or increase or decrease of a signal generated from the inhibiting activity, wherein the absence or decrease of the signal is indicative for a putative inhibitor.
22. The method of claim 20 or 21, wherein said cell is a cell of claim 9, obtained by the method of claim 10 or can be obtained by the transgenic non-human animal of claim 15 or 16.
23. A method of identifying and obtaining a pro-drug or drug capable of modulating the activity of a molecular variant of a CYP3A5 polypeptide comprising the steps of:
- (a) contacting the host cell of claim 8, the cell obtained by the method of claim 10, the polypeptide of claim 11 or the solid support of claim 17 or 18 with the first molecule known to be bound by a CYP3A5 polypeptide to form a first complex of said polypeptide and said first molecule;
  - (b) contacting said first complex with a compound to be screened, and
  - (c) measuring whether said compound displaces said first molecule from said first complex.

24. A method of identifying and obtaining an inhibitor capable of the activity of a molecular variant of a CYP3A5 polypeptide or its gene product comprising the steps of:
  - (a) contacting the host cell of claim 8, the cell obtained by the method of claim 10, the protein of claim 11 or the solid support of claim 17 or 18 with the first molecule known to be bound by a CYP3A5 polypeptide to form a first complex of said protein and said first molecule;
  - (b) contacting said first complex with a compound to be screened, and
  - (c) measuring whether said compound displaces said first molecule from said first complex.
25. The method of claim 23 or 24, wherein said measuring step comprises measuring the formation of a second complex of said protein and said compound.
26. The method of any one of claim 23 to 25, wherein said measuring step comprises measuring the amount of said first molecule that is not bound to said protein.
27. The method of any one of claims 23 to 26, wherein said first molecule is labeled.
28. A method for the production of a pharmaceutical composition comprising the steps of the method of any one of claims 20 to 27; and the further step of formulating the compound identified and obtained or a derivative thereof in a pharmaceutically acceptable form.
29. A method of diagnosing a disorder related to the presence of a molecular variant of a CYP3A5 gene or susceptibility to such a disorder comprising determining the presence of a polynucleotide of any one of claims 1 to 3 or the gene of claim 4 or 5 in a sample from a subject.
30. The method of claim 29 further comprising determining the presence of a polypeptide of claim 11 or the antibody of any one of claims 12 to 14.

31. A method of diagnosing a disorder related to the presence of a molecular variant of a CYP3A5 gene or susceptibility to such a disorder comprising determining the presence of a polypeptide of claim 11 or the antibody of any one of claims 12 to 14 in a sample from a subject.
32. The method of any one of claims 29 to 31, wherein said disorder is cancer or diseases including cardiovascular diseases, diabetes and AIDS.
33. The method of any one of claims 29 to 32 comprising PCR, ligase chain reaction, restriction digestion, direct sequencing, nucleic acid amplification techniques, hybridization techniques mass spectroscopy or immunoassays.
34. A method of detection of the polynucleotide of any one of claims 1 to 3 or the gene of claim 4 or 5 in a sample comprising the steps of
  - (a) contacting the solid support of claim 17 or 18 with the sample under conditions allowing interaction of the polynucleotide of claim 1 to 3 or the gene of claim 4 or 5 with the immobilized targets on a solid support and;
  - (b) determining the binding of said polynucleotide or said gene to said immobilized targets on a solid support.
35. An in vitro method for diagnosing a disease comprising the steps of the method of claim 34, wherein binding of said polynucleotide or gene to said immobilized targets on said solid support is indicative for the presence or the absence of said disease or a prevalence for said disease.
36. A diagnostic composition comprising the polynucleotide of any one of claims 1 to 4, the gene of claim 3 or 4, the vector of claim 6 or 7, the polypeptide of claim 11 or the antibody of claim 12 or 13.
37. A pharmaceutical composition comprising the polynucleotide of any one of claims 1 to 3, the gene of claim 4 or 5, the vector of claim 6 or 7, the polypeptide of claim 11 or the antibody of claim 12 or 13.

38. Use of the polynucleotide of any one of claims 1 to 3, a polynucleotide comprising SEQ ID No: 104, a polynucleotide encoding a polypeptide comprising SEQ ID No: 145, the gene of claim 4 or 5, the vector of claim 6 or 7, the polypeptide of claim 11, a polypeptide comprising SEQ ID No: 145 or the antibody of claim 12 or 13 for the preparation of a diagnostic composition for diagnosing a disease.
39. Use of the polynucleotide of any one of claims 1 to 3, a polynucleotide comprising SEQ ID No: 104, a polynucleotide encoding a polypeptide comprising SEQ ID No: 145, the gene of claim 4 or 5, the vector of claim 6 or 7, the polypeptide of claim 11, a polypeptide comprising SEQ ID No: 145 or the antibody of claim 12 or 13 for the preparation of a pharmaceutical composition for treating a disease.
40. Use of a polynucleotide selected from the group consisting of:
  - (a) a polynucleotide having the nucleic acid sequence of SEQ ID NO: 82, 88, 104 or 112;
  - (b) a polynucleotide encoding a polypeptide having the amino acid sequence of SEQ ID No: 127, 132, 141 or 145;
  - (c) a polynucleotide capable of hybridizing to a CYP3A5 gene, wherein said polynucleotide is having at least one additional nucleotide at a position corresponding to position 3709/3710 or 27131/27132 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614) or a nucleotide exchange at a position corresponding to position 7303 or 27289 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614);
  - (d) a polynucleotide capable of hybridizing to a CYP3A5 gene, wherein said polynucleotide is having an additional G nucleotide at a position

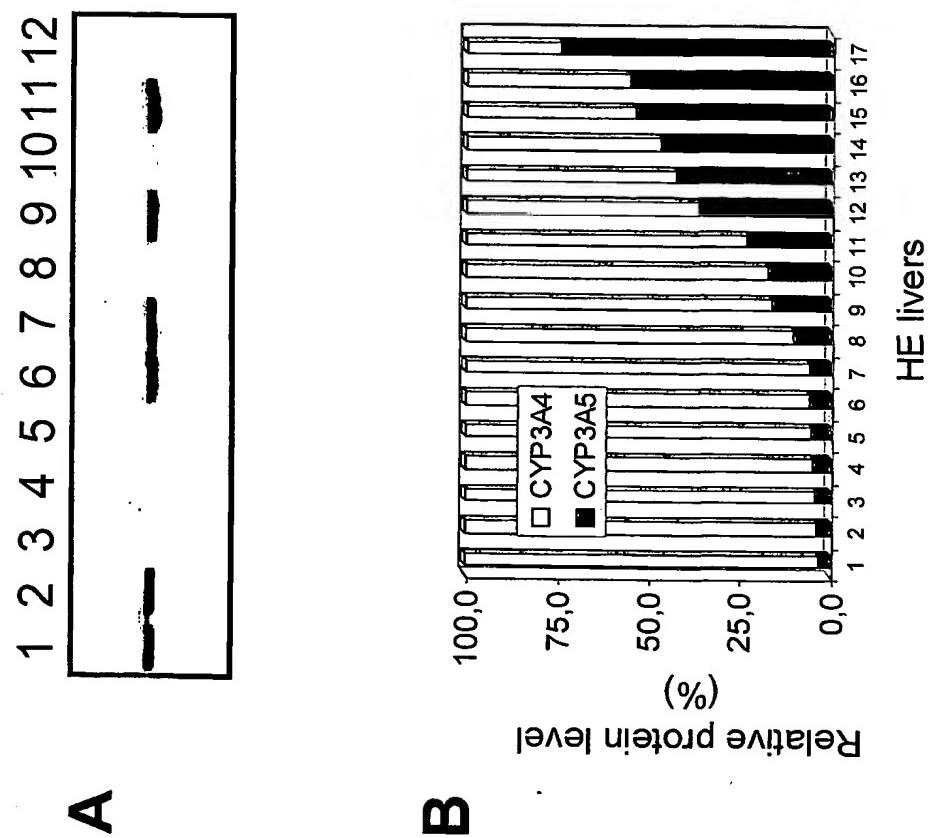
corresponding to position 3709/3710 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614), an additional T nucleotide at a position corresponding to position 27131/27132 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614), or an A at a position corresponding to position 7303 or 27289 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614);

for the preparation of a diagnostic composition for diagnosing a disease in a subject having a genome comprising a variant allele of the CYP3A5 gene, wherein said allele is having an A at a position corresponding to position 6986 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614).

41. Use of a polynucleotide comprising a polynucleotide having an A at a position corresponding to position 14690 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614) for the preparation of a diagnostic composition for diagnosing a disease in a subject having a genome comprising a variant allele of the CYP3A5 gene, wherein said allele is having an A at a position corresponding to position 6986 of the CYP3A5 gene (Accession No: AF280107.1, wherein position 166220 has been numbered +1 and position 174832 has been numbered +8613, and Accession No: AC005020.2, wherein position 27341 has been numbered +8614).

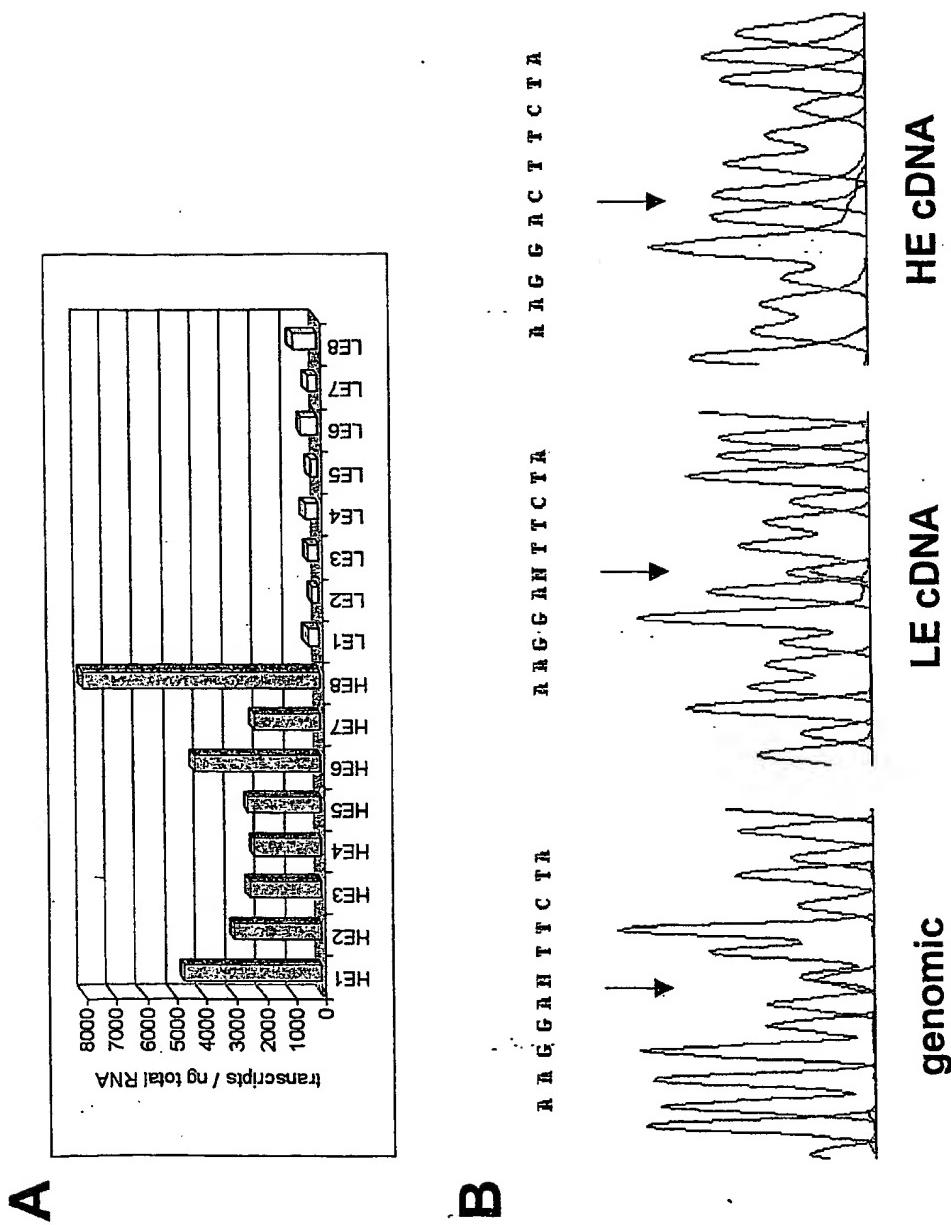
42. The use of claim 40 or 41, wherein said subject is an African American.
43. The use of any one of claims 38 to 41, wherein said disease is cancer or diseases including cardiovascular diseases, diabetes and AIDS.
44. A diagnostic kit for detection of a single nucleotide polymorphism comprising the polynucleotide of any one of claims 1 to 3, the gene of claim 4 or 5, the vector of claim 6 or 7, the polypeptide of claim 11, the antibody of claim 12 or 13, the host cell of claim 8, the transgenic non-human animal of claim 15 or 16 or the solid support of claim 17 or 18.

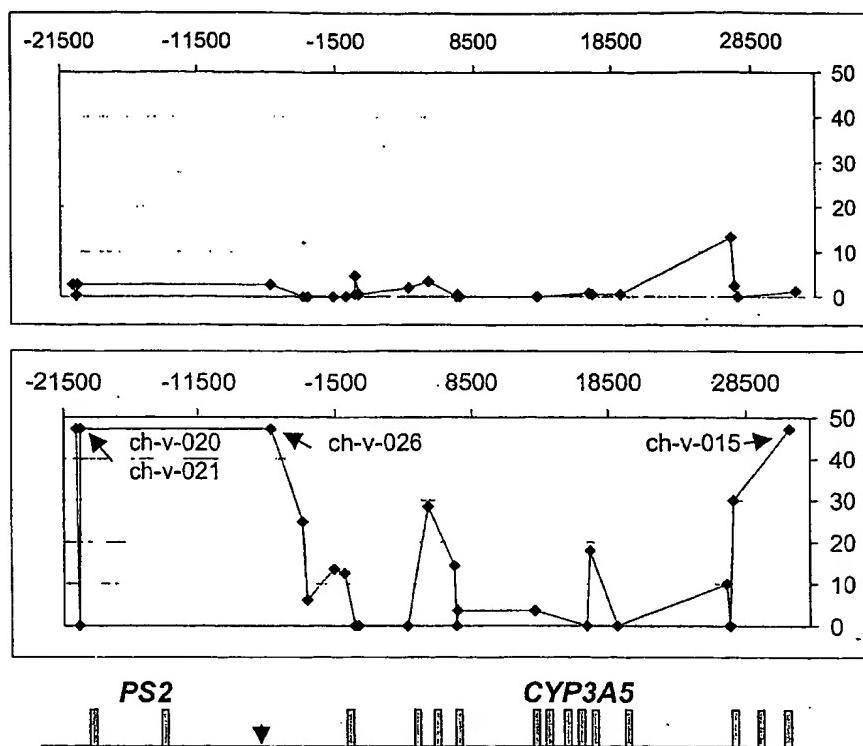
Fig. 1



2/21

**Figure 2**



**Figure 3**

4/21

## Figure 4

Seq ID: 113

>chzk\_ch-v-020\_254T>G\_ch-v-021\_582A>G

```

1 tggtcaccca ccatgtgtac agtaccctgc tagggtccag ggtcatgaaa gtaaataata
61 ccagactgtg cccttgagga actcacctct gctaaggaa acaggcacag aaacccacaa
121 gggtgtaga gaggaaatag gacaatagga ctgtgtgagg gggataggag gcacccagag
181 gaggaaatgg ttacatctgt gtgaggaggt tggtaaggaa agacttaat agaagggtc
241 tgtctggctg ggctgtcaag gatgtgttagg agtcatctag ggggcacaag tacactccag
301 gcagagggaa ttgcatgggt aaagatctgc agttgtggct tggggatg gatttcaagt
361 attctggaat gaagacagcc atggaaacaa gggcaggtga gaggatattt aagaggctc
421 atgccaatgg ctccacttca gtttctgata agaactcagg tccctggac tccctgataa
481 aactgattaa gttgtttatg attccccata gaatatgaac tcaaaggagg taagcaaagg
541 ggtgtgtcg attctttgct actggctgca gctgcagccc caccccttc tccagcacat
601 aaacatttca gcagcttgac ctaagactgc tggcagggc agggatgctc caggcagaca
661 gcccagcaaa caacagcaca cagctgaaag taagactcag aggagacagt tgaagaaggc
721 aagtggcgat ggacctcatc ccaaatttgg cggtggaaac ctggcttctc ctggctgtca
781 gcctggtgct cctctatctg tcagtaactg tccagattcc tctcctctgt

```

Seq ID: 114

>chzk\_ch-v-031\_318G>A

```

1 tggtcaccca ccatgtgtac agtaccctgc tagggtccag ggtcatgaaa gtaaataata
61 ccagactgtg cccttgagga actcacctct gctaaggaa acaggcacag aaacccacaa
121 gggtgtaga gaggaaatag gacaatagga ctgtgtgagg gggataggag gcacccagag
181 gaggaaatgg ttacatctgt gtgaggaggt tggtaaggaa agacttaat agaagggtc
241 tgtctggctg ggctgtcaag gatgtgttagg agtcatctag ggggcacaag tacactccag
301 gcagagggaa ttgcatgagt aaagatctgc agttgtggct tggggatg gatttcaagt
361 attctggaat gaagacagcc atggaaacaa gggcaggtga gaggatattt aagaggctc
421 atgccaatgg ctccacttca gtttctgata agaactcagg tccctggac tccctgataa
481 aactgattaa gttgtttatg attccccata gaatatgaac tcaaaggagg taagcaaagg
541 ggtgtgtcg attctttgct actggctgca gctgcagccc caccccttc tccagcacat
601 aaacatttca gcagcttgac ctaagactgc tggcagggc agggatgctc caggcagaca
661 gcccagcaaa caacagcaca cagctgaaag taagactcag aggagacagt tgaagaaggc
721 aagtggcgat ggacctcatc ccaaatttgg cggtggaaac ctggcttctc ctggctgtca
781 gcctggtgct cctctatctg tcagtaactg tccagattcc tctcctctgt

```

Seq ID: 115

>chzk\_ch-v-032\_544G>A

```

1 tggtcaccca ccatgtgtac agtaccctgc tagggtccag ggtcatgaaa gtaaataata
61 ccagactgtg cccttgagga actcacctct gctaaggaa acaggcacag aaacccacaa
121 gggtgtaga gaggaaatag gacaatagga ctgtgtgagg gggataggag gcacccagag
181 gaggaaatgg ttacatctgt gtgaggaggt tggtaaggaa agacttaat agaagggtc
241 tgtctggctg ggctgtcaag gatgtgttagg agtcatctag ggggcacaag tacactccag
301 gcagagggaa ttgcatgggt aaagatctgc agttgtggct tggggatg gatttcaagt
361 attctggaat gaagacagcc atggaaacaa gggcaggtga gaggatattt aagaggctc
421 atgccaatgg ctccacttca gtttctgata agaactcagg tccctggac tccctgataa
481 aactgattaa gttgtttatg attccccata gaatatgaac tcaaaggagg taagcaaagg
541 ggtatgtcg attctttgct actggctgca gctgcagccc caccccttc tccagcacat
601 aaacatttca gcagcttgac ctaagactgc tggcagggc agggatgctc caggcagaca
661 gcccagcaaa caacagcaca cagctgaaag taagactcag aggagacagt tgaagaaggc
721 aagtggcgat ggacctcatc ccaaatttgg cggtggaaac ctggcttctc ctggctgtca
781 gcctggtgct cctctatctg tcagtaactg tccagattcc tctcctctgt

```

5/21

### **Figure 4 continued**

Seq ID: 116

>chzk ch-v-033 550G>A

1 tggtcaccca ccatgtgtac agtaccctgc tagggtccag ggtcatgaaa gtaaaaataata  
61 ccagactgtg cccttgagga actcacctct gctaaggaa acagggcacag aaaccccacaa  
121 gggtgttaga gaggaaatag gacaatagga ctgtgtgagg gggataggag gcacccagag  
181 gaggaaatgg ttacatctgt gtgaggaggt tggtaaaggaa agactttaat agaagggggt  
241 tgctctggctg ggcttgcaga gatgtgtagg agtcatctag ggggcacaag tacactccag  
301 gcagagggaa ttgcattgggtaaagatctgc agttgtggct tggggatg gatttcaagt  
361 attctggaat gaagacagcc atggaaacaa gggcaggtga gaggatattt aagaggcttc  
421 atgccaatgg ctccacttca gtttctgata agaactcagg ttccgtggac tccctgataaa  
481 aactgattaa gttgtttatg attccccata gaatatgaac tcaaaggagg taagcaaagg  
541 ggtgtgtgca attctttgc actggctgca gctgcagccc caccccttc tccagcacat  
601 aaacatttca gcagcttgac ctaagactgc tggcagggc agggatgctc caggcagaca  
661 gcccagcaaa caacagcaca cagctgaaag taagactcag aggagacagt tgaagaaggc  
721 aagtggcgtat ggacacctatc ccaaatttgg cggtggaaac ctggcttctc ctggctgtca  
781 gcctgggtgtc cctctatctg tcagtaactg tccagattcc tctctctgtt

Seq ID: 117

>chz1 ch=v=026 229A>G

```
1 ggagtgcacct gatTTTccag gtgtgtctg tcacCCCTT ctttgactag gaaagggaac  
61 tccttggacc cttgcgcTTc tcaagtgagg caatgcCTcg ccctgcTTcg gcttggcac  
121 agcacgcTgc acccactgTC ctgcacCCAC tgcTGGCAC tccCTAGTGA gatgaACCCG  
181 gtacCTcaga tgaaaATGCA gaaatCACCC gtCTTCTGTg tcactcAC tc tgggagCTGT  
241 agacCCggAGC tggTCCTATT CGGCCATCTT ggCTCCACCC CCGAGTTT ggCTTTTAAT  
301 tgaaAGTgTA ttgatgtggg aaggagataa tgccatgcat ttatgagcac atattagagg  
361 gtctgagaca atgcATgtGA taaaAGGTCA CCTAAGGAAG aaaaaAAAGAA caAGGGAAAG  
421 cactggaaAG aacgtgtGc tgggAGtccc tgggCCACCA aagtCTGGAG aaaaAGTgta  
481 accacaAGGC tcccAGCCTA gtttcaCTGA agacCTCGAC actAGGTGAC ttatGGGATC  
541 ctTggtagga caagCCTGA aaAGTTTAC aatAGCAAT gtggACGTG tcagaACCAA  
601 atgatgtcac gtgtgttattt gtgtgtgtgt gtcagtgtgt gtgttAAAAA atcatgacAA  
661 ataaaAGCAGG ctgtGAAGAG gggattCCCA tgctcgtgtg cctgataACA caactatCAC  
721 aaacgCTTT CGAAACCCACA agtttgcaca aaggcaatCC caacCtttACA caaaaa
```

Seq ID: 118

>chzi ch-v=027 566G>A

1 taatggaaaat aagaatttatt ttgatggctc taacagtgc ac atttatataca tctgtttt  
61 ctggaggcatt ctataataag ttatattaa gcaaataatca aaaaacccat tacaaaagta  
121 tcatcgata ctttcctgaa cattaaggag aaatctatag aactgaatga atgagaaccca  
181 acaaaggtaaat atatgtgatc attgtaacca ttgttgggtg gggcatttg tcagaactcc  
241 aatgtgatta ttaacatagg tgagaattaa tccactgtg ctttgcctat tgcttagaaa  
301 gaacattcat agtttaatta tgccctttt gaccaagcac agtgcctat gcctgttaatc  
361 ccagcacttt gggaggccga ggtgggtgga tcacctgagg tcaggagttc gagaccagcc  
421 tgaccaacat ggtgaaaaccc catctctact aaaaatacaa aaattagcta ggtgtgggg  
481 tatgcaccta taatctcagc taccaggag gctgaggcag gagaatact tgaacctgga  
541 ggcagagggtt gcagtgagcc gagaca<sup>a</sup>cacc attacactcc agcctgggtg acagagtgg  
601 attccatctc aaaaaaaaaa aaaaaaaaaatt atgcctttt gaagcacata cattttataaa  
661 catacaactg aatcccttat tatattatta gttttgattt aatgtttca aaccatctcc  
721 cctgatattt ctggggatg ggaaacatgt ttcttacac ctcttcattt ccattctcaa  
781 ctccccactg tcttactgca atgaacactt aataagaaaac agtcaattgg tcaattgg  
841 gggcaacacagg ctaaacacac tcattccttgc tctgttccca cttctttctt tactttccct  
901 tcctgagtaa cttatcttaa agtcattagg tgggtggcag ccagatggtg gcccacacatt  
961 aaggtagaaaa agagagtgtc atgatggttc caagtcagag accttagtgg gtgaggatca  
1021 agtaggtgtt cacgtggaga aacagccccgg cctgtgtgtg ggagtccaag caagcagaga  
1081 aaatgtcgac acagagggtt ggctqaaaa

6/21

## Figure 4 continued

Seq ID: 119

&gt;chzh\_ch-v-028\_601G&gt;A

```

1 caaaaattag ctagggtgg tggatgcac ctataatctc agtacccag gaggctgagg
61 caggagaatc acttgaacct ggaggcagag gttgcagtga gccgagacgc accattacac
121 tccagcctgg gtgacagagt gagattccat ctcaaaaaaaaaa attatgcctt
181 tttaagec acatattta taacatacaa ctgaatccc tattatatta ttagtttga
241 tttaatgtt tcaaaccatc tcccctgata tttctggag atggaaaca tggtttctta
301 cacctcttc attccattct caactccaa ctgtcttact gcaatgaaca ctaataaga
361 aacagtcaat tggtcaattt gttggcaac aggctaaaca cactcattcc ttgtctgttc
421 ccacttctt cttaacttcc ctttcctgag taacttatcc taaagtctt aggtgggtgg
481 cagccagatg gtggccacac attaaggttag aaaaagagat gtcatgatgg ttccaagtca
541 gagacctagt agggtgagga tcaagtaggt gttcacgtgg agaaaacagcc cggcctgtgt
601 atgggagtcc aagcaagcg agaaaaatgtc gacacagagg ggtggccgtga aaaaagcagcc
661 agagcctaaa cagggcatgg agaacatatt tagggcatga ggtgaggagg gcatccatga
721 gtgggaaggg atgggtgagg tttcactaca taaagggat ttagtggaaaata agtaataaaa
781 gtatactggaa agccaggtgt gtcactttt cagaaaaagag tcattggattc agaaaagg

```

Seq ID: 120

&gt;chzg\_ch-v-029\_464T&gt;C

```

1 ttctttctt acttccctt cctgagtaac ttatcctaaa gtcatttaggt gggtggcagc
61 cagatggtgg ccacacatta aggtagaaaa gagagtgtca tgatggttcc aagtcagaga
121 cctagtaggg tgaggatcaa gtaggtgttc acgtggagaa acagccccgc ctgtgtgtgg
181 gagtccaagc aaggcagagaa aatgtcgaca cagaggggtg gcctgaaaaa gcagccagag
241 cctaaacagg gcatggagaa catattnn gcatgaggtg aggagggcat ccatgagtgg
301 gaagggatgg gtgagggttc actacataaa ggggattgtat gaaataagta aataaaagtat
361 actggaagcc aggtgtgtca cttttcggaa aaagagtcat ggattcgaaa agggagaaaa
421 ctagcagggaa tcctatgaaa ttagattaaa atggatgtat ccagttatat tcataccctt
481 ctagatagat aaatgggttag atagggtgata aaaagataac aagaggacaa gataattaga
541 tagacataaa tgtatgtatg tttttgtgtg tgtgtacaaa aaaacatata ctccctactt
601 ctctccactg atagggcttag gtaacaatgg catttcata gcaatgagca cacttagtgg
661 ccagatctt gcttataat accatttcc actgaaaagga accagagctt ttttagagaaaa
721 tggctgattc cagggccagg attaagaatg ttcaagataa gcctaggata cattttgtgc
781 caggaagcaa gaagatgttc aaatgatttca agtataatgt ttggaaatga tatttggaaaa
841 tgatttccaa atgatatttca caaatgattt ccaaataatgata tatggaaaca cttaaaagact
901 ccactaaaga actatttagat ctgataaaaca aattcagttaa ttgtgtctgg tacaaaatca
961 acatacaaaa accagtagca ttctgcattt ccaacagtga acaatctggc aaaaataaaaa
1021 aatgtatcc catttacaat aaccccaat aaaactaaat acctggaaat taacttaaga
1081 gaaagatgtc tacaattat attgtaaaac actgtatggaa gaaattgtt aagacacaaaa
1141 aaagaaggat attccatgtt tatattttt aagcattaaat attgtttaaaa atgtcca

```

Seq ID: 121

&gt;chza\_ch-v-034\_328T&gt;C

```

1 caaagcaaaa atggacaaaat ggatcagatc aagttaaaaaa gcttctgtac cacaagaaaa
61 gcaatcaaca aagtgaagac acaaaccaca gaatgggaga aaatattttc aaagtccacac
121 tctgacaaca gattaatgc cagaatacat gaagcgtca aacaactctg taaggaaaaa
181 tctaataatc caatcaaaaa atgggcaaaa tttgaataga catttttcaa aagaagacat
241 acaaataatccca cataggcata tgataagggtc ctcaacatca ctggcattt gagaatgtca
301 aatcaaaaacc acaatggat atcatctac cccagctaaa atggtttttca tccaaaagac
361 aggcaacaaac aaatgccgc gagaatgtgg agaaaaggga acccttgtac actgttggtg
421 taaattatgtt caaccactat agagaacaaat ttggaggttc ctcaaaaacat taaaattaac
481 attaaataga gctaccacaa tatccagaaaa tccccatgtct ggttatatac ctggaaagaaa
541 gggaaatcata tattgaagag ataacatcac tccaatattc acaatagccca ctattcacaa
601 atgccaagat ttggaaagcaa ctaaagtgtc catcaacaga tgaatggata aagaaagtac
661 tccaattata cacaatggag cacaatttc ccatgaaaaa agcatgagat cctgttatct
721 gtaataatataat ggtatggact ggaggtcattc atgttaatgtt aataaagccca ggcacagaaaa
781 cacagatattt gcaagtttcc acataacttgc gggatct

```

7/21

**Figure 4 continued**

Seq ID: 122

&gt;chzd\_ch-v-030\_683T&gt;A

```

1 cattaaaatt aacataaaat agagctacca caatatccag aaatccccat gctgggtata
61 tacctggaaag aaaggaaatc atatattgaa gagataacat cactccaata ttcacaata tag
121 ccactattca caaatgccaa gatttggaaag caacctaagt gtccatcaac agatgaatgg
181 ataaaagaaaag tactccaatt atacacaatg ggcacaatt cagccatgaa aaaagcatga
241 gatcctgtta tctgtataaa tatggatgga actggaggc atcatgttaa gtgaaataag
301 ccaggcacag aaacacagat attgcaagtt ctcacatact tttggatct acaaataaaa
361 acaactgagc taatgtctgg gccttagtca gtgtgttacc caagtactgg gggcacagct
421 tttaaaaatac atcatgaatg cttaatacaca ggaatgaata gatgagagggc acaaactgg
481 tgggtgttct tctgtatacac agtatcttcc ttgacagatt cagttacaact ctcacacgg
541 aagtctcttc atgttatgtt accttatgag gaattaagtgc gggcacatg atttcttata
601 ttttcctttc cagaacaaga ccaactttat tagtggac acagtgtggc tgcatttgag
661 tcccaagcaa ccatttgtct atagctatca ccacagagtc agaggggatg agaccccag
721 caatctcacc caagacaact ccaccaacat tcctggttac ccaccatgtg tacagtaccc
781 tgcttaggaac cagggtcatg aaaaataata accagact gtgccttga ggagctcacc
841 tctgctaagg gaaacaggca tagaaactta caatgggtt agagagaaaa gggacaata
901 ggactgtgtg agggggatag gggcaccca gaggaggaaa tggttacatt tttgtgagga
961 ggttggtaag gaaaaattttt agcagaagggtt gctgtctgg ctgggcttgg aaggatatacg
1021 aggagtcatc tagagggcac aggtacactc caggcagagg gaatttcgtt ggtaaagatg
1081 tgttaggtgtg gcttgtgagg atggatttca attattcttag aatgaa

```

Seq ID: 123

&gt;chzy\_ch-v-002\_159G&gt;A

```

1 gattaagctt ttcatgattc ctcatagaac atgaactcaa aagaggtcag caaaggggtg
61 tttgcgattc tttgttattt gctgcagcta tagccctgcc tccttctcca gcacataaaat
121 ctttcagcag cttggctgaa gactgtgtg cagggcagg aagctccagg caaacagccc
181 agcaaaacagc agcaactcagc taaaagggaaag actcacagaa cacagttgaa gaaggaaagt
241 ggcgATGGAC CTCATCCCAA ATTTGGCGGT GGAAACCTGG CTTCTCCTGG CTGTCAGCCT
301 GGTGCTCCTC TATCTgttag taactgtcca aactcccttc ttgtttctt tggacttggg
361 gtgctaattcg ggccctttt cccttatctg ttttgaagat caaaagagat gttcaaggag
421 aagtagctga agtggtagac gctacaaacg catagaagtt attattatct tatgcagatc
481 tatgaatgaa taaataagca tttctccat ccacccctta attttgggtga cttaggaggg
541 ttagggacacg cattttgttag tgggaatgtt ttgatttagct tagatctgac gaagactaat
601 caatgaaaac atggcagcgg caga

```

Seq ID: 124

&gt;chzy\_ch-v-003\_171C&gt;T

```

1 gattaagctt ttcatgattc ctcatagaac atgaactcaa aagaggtcag caaaggggtg
61 tttgcgattc tttgttattt gctgcagcta tagccctgcc tccttctcca gcacataaaat
121 ctttcagcag cttggctgaa gactgtgtg cagggcagg aagctccagg caaacagccc
181 agcaaaacagc agcaactcagc taaaagggaaag actcacagaa cacagttgaa gaaggaaagt
241 ggcgATGGAC CTCATCCCAA ATTTGGCGGT GGAAACCTGG CTTCTCCTGG CTGTCAGCCT
301 GGTGCTCCTC TATCTgttag taactgtcca aactcccttc ttgtttctt tggacttggg
361 gtgctaattcg ggccctttt cccttatctg ttttgaagat caaaagagat gttcaaggag
421 aagtagctga agtggtagac gctacaaacg catagaagtt attattatct tatgcagatc
481 tatgaatgaa taaataagca tttctccat ccacccctta attttgggtga cttaggaggg
541 ttagggacacg cattttgttag tgggaatgtt ttgatttagct tagatctgac gaagactaat
601 caatgaaaac atggcagcgg caga

```

8/21

**Figure 4 continued**

Seq ID: 125

&gt;chzy\_ch-v-004\_418-420delGAG

```

1 gattaagctt ttcatgattc ctcatagaac atgaactcaa aagaggtcag caaagggttg
61 tgtgcattc tttgttattt gctgcagcta tagccctgcc tccctctcca gcacataaat
121 ctttagcag cttggctgaa gactgctgtg cagggcaggg aagctccagg caaacagccc
181 agcaaacagc agcaactcagc taaaaggaag actcacagaa cacagttgaa gaaggaaagt
241 ggcgATGGAC CTCATCCCAA ATTTGGCGGT GGAAACCTGG CTTCTCCTGG CTGTCAGCCT
301 GGTGCTCCTC TATCTgtgag taactgtcca aactcccttc tttgtttcct tggacttggg
361 gtgctaatcg ggcgcctttt cccttatctg ttttgaagat caaaaagagat gttcaag||aag
421 tagctgaagt gttggacgct acaaacgcattt agaagtttattt attatcttat gcagatctat
481 gaatgaataa ataagcattt cuccatcca ctttctaatt ttggtagacta ggagggtta
541 gggacagcat ttggtagtgg gaatgattttag attagcttag atctgacgaa gactaatcaa
601 tggaaacatg gcagcggcag a

```

Seq ID: 126

&gt;chzx\_ch-v-005\_187C&gt;T\_ch-v-006\_191-192insG

```

1 agggaaaggac ctgatgagtg aatgcaatta ctgatgttgg agttgctgtt attattttatc
61 gtgtacatat taccccttc tcttgaccat tccagttcct gagtaactca ccagccctct
121 gatctataaa gtcacaatcc ctgtgacctg atttctgttt cactttgttag ATATGGGACC
181 CGTACA[TATG] G[GACTTTTA AGAGACTGGG AATTCCAGGG CCCACACCTC TGCCTTGTT
241 GGGAAATGTT TTGTCCTATC GTCAAGgtgag ttgcttggc ttccctttt gcttctttag
301 gttgcaaaaca tcagcttagt tccatcagta aaaatgcctt tccttggag ggagttctga
361 ggtttcacat ttccagaaat ggtgggactg ggtgcagtgg atcatgcctg taatctcagc
421 ctctgtgagg ccaagactgg caaattgtt gagcccagga gtttg

```

Seq ID: 127

&gt;peptide of ch-v-005\_ch-v-006

1 MDLIPNLAVE TWLLLAVSLV LLYLYGTRTY GTF.

Seq ID: 128

&gt;chzw\_ch-v-007\_143C&gt;T

```

1 catagacaag ggtgagtcct tcagttactta gagaaaattc aagagtgact ttaaattcccc
61 cacttcaaat atattctctg tttcttgc tttcccttaa gacatctctg aatagcttcc
121 ttcaactgcc agtggaaat aggggcctg atttcattgg acgcaactgt tttcagcccc
181 aatttagaggtt agggttttt ctataaaa taataatcaa ctgttatttt gtttcctctc
241 ccagGGTCTC TGGAAATTG ACACAGAGTG CTATAAAAAG TATGGAAAAA TGTGGGGGtg
301 agtattctga aaacccat tggatagacc tgctactgtg aggaggttac cccactgcag
361 gatagtctct gcccaggctc tcatggatg aagctttgt caacctaata acaaacagag
421 agaggcttc tggaaagaaga ggataattac ttgggagtag aatattgcaa tggaaatctg
481 cttgcgtta taaaactatgt gcaattttag ggaggtaaaac aagacaaga tgctccatag
541 aaaaatatgag aagaatctca taactgtttt gagataatta ttgttagcta caaagatcaa
601 ta

```

Seq ID: 129

&gt;chzv\_ch-v-008\_199C&gt;A

```

1 cagtatctt cccctttt gaccacatta cccttcattca tatgaaggct tgggtggctc
61 ctgtgtgaga ctcttgcgtgt gtgtcacacc ctaatgaact agaacctaag gttgctgtgt
121 gtcgtaac taggggtatg gattacataa cataatgatc aaagtctggc ttccctgggtg
181 tggctccagc tgcagaattg ggctagtggaa gtttaatcag ctccgttgc cccacacagA
241 ACGTATGAAG GTCAACTCCC TGTGCTGGCC ATCACAGATC CCGACGTGAT CAGAACAGTG
301 CTAGTGAAAG AATGTTATTG TGTCTTCACA AATCGAAGGg taagcatcca ttttttggaaa
361 tttaaataat gattgtatcca ctgattaaat ttttatttttgg aaaaaaaaaac atattcacag
421 aaggttacct aaaaaatgtt cagggagttt ccatgtactc ttccatctgt cccgccccagt
481 ggttaacatct tgcaatcttgc tatattgcaa tatataatcta gtatattcat attatcaggt
541 tggcacaaaa gttaaaaatgg caaactacag gctggcata atggctcatg cctg

```

9/21  
Figure 4 continued

Seq ID: 130

&gt;chzv\_ch-v-025\_224C&gt;T

```

1 cagtatctt tccctgttt gaccacatta cccttcata tatgaaggct tgggtggc
61 ctgtgtgaga ctcttgctgt gtgtcacacc ctaatgaact agaacctaag gttgctgtgt
121 gtcgtacaac taggggtatg gattacataa cataatgatc aaagtctggc ttccctgggtg
181 tggctccagc tgcagaatcg ggcttagtcaa gtttaatcag ctctgttgc cccacacagA
241 ACGTATGAAG GTCAACTCCC TGTGCTGGCC ATCACAGATC CCGACGTGAT CAGAACAGTG
301 CTAGTGAAG AATGTTATTTC TGTCTTCACA AATCGAAGGg taagcatcca ttttttggaa
361 tttaaataat gattgatcca ctgattaaat ttttattttg aaaaaaacat atattcacag
421 aaggttacct aaaaatgtt caggaaggtt ccatgtactc ttcatctgt cccggcccagt
481 ggtaacatct tgcaatctt gatatattgaa tatataatcta gtatattcat attatcaggt
541 tggcacaaaaa gttttttttt gttttttttt caaaactacag gctggcata atggctcatg cctg

```

Seq ID: 131

&gt;chzv\_ch-v-009\_320C&gt;A

```

1 cagtatctt tccctgttt gaccacatta cccttcata tatgaaggct tgggtggc
61 ctgtgtgaga ctcttgctgt gtgtcacacc ctaatgaact agaacctaag gttgctgtgt
121 gtcgtacaac taggggtatg gattacataa cataatgatc aaagtctggc ttccctgggtg
181 tggctccagc tgcagaatcg ggcttagtcaa gtttaatcag ctccgttgc cccacacagA
241 ACGTATGAAG GTCAACTCCC TGTGCTGGCC ATCACAGATC CCGACGTGAT CAGAACAGTG
301 CTAGTGAAG AATGTTATTTC TGTCTTCACA AATCGAAGGg taagcatcca ttttttggaa
361 tttaaataat gattgatcca ctgattaaat ttttattttg aaaaaaacat atattcacag
421 aaggttacct aaaaatgtt caggaaggtt ccatgtactc ttcatctgt cccggcccagt
481 ggtaacatct tgcaatctt gatatattgaa tatataatcta gtatattcat attatcaggt
541 tggcacaaaaa gttttttttt gttttttttt caaaactacag gctggcata atggctcatg cctg

```

Seq ID: 132

&gt;peptide of ch-v-009\_320C&gt;A

```

1 MWGTYEGQLP VLAITDPDVI RTVLVKECYV VFTNRRSLGP VGFMKSAISL AEDEEWKRIR

```

Seq ID: 133

&gt;chzv\_ch-v-018\_441-444insCTAAAAAAAT

```

1 cagtatctt tccctgttt gaccacatta cccttcata tatgaaggct tgggtggc
61 ctgtgtgaga ctcttgctgt gtgtcacacc ctaatgaact agaacctaag gttgctgtgt
121 gtcgtacaac taggggtatg gattacataa cataatgatc aaagtctggc ttccctgggtg
181 tggctccagc tgcagaatcg ggcttagtcaa gtttaatcag ctccgttgc cccacacagA
241 ACGTATGAAG GTCAACTCCC TGTGCTGGCC ATCACAGATC CCGACGTGAT CAGAACAGTG
301 CTAGTGAAG AATGTTATTTC TGTCTTCACA AATCGAAGGg taagcatcca ttttttggaa
361 tttaaataat gattgatcca ctgattaaat ttttattttg aaaaaaacat atattcacag
421 aaggttacct aaaaatgtt ctaaaaaat gaaggttcca tgactcttc atccgttcc
481 gccccgtgtt aacatcttgc aatcttgcattt attgcaat atatctgtt tattcatatt
541 atcaggttgg cacaaaatgtt aaaaatggcaa actacaggct gggcataatg gctcatgcct
601 g

```

Seq ID: 134

&gt;chzv\_ch-v-016\_145T&gt;G

```

1 agcggaaaaac tcaaggaggt atgaaaataa gatgagtctt aatttagaaat gtaaagaatg
61 aatctgggga caggtagaaa gtaagatcac agtccgttcc caagggttag tccactgagt
121 tgcagctttcc taaaaatgtt ctttgatttt tatgtacaga aaagacatca caaaattcat
181 tacaaaatgtt cacttactgc tccatgctgg agaaaaggcat atccttctgg gacttgagt
241 tgcacattt actacaggta ctgatctgtt ttgtgcttag ATGTTCCCCA TCATTGCCCA
301 GTATGGAGAT GTATTGGTGA GAAACTTGAG GCGGGAAAGCA GAGAAAGGCA AGCCTGTCAC
361 CTTGAAAGAg taagttaggag cacagccatg gggttctgag ctgtcatgag cccttccagc
421 tgcctgcatg ggagtcgaca gtcgcactgt tgggttactc cagtgaccag aaaaaagcag
481 ggccagcgcgtt caactccaaa gagccaccta agagggtt gctccatga ggcggcaagt
541 cagcaaggga aaaggccctt ctctcctgtt cacaggagcc aggattact tatctgttaa
601 ctt

```

10/21  
Figure 4 continued

Seq ID: 135

>chzt\_ch-v-019\_241T>C

```

1  agcggaaaac tcaaggaggt atgaaaataa gatgagtctt aattagaaat gtaaagaatg
61  aatctgggaa caggtagaaa gtaagatcac agtccgttc caagggtag tccactgagt
121 tgcagcttcc taaaaatggt cttttatctt tatgtacaga aaagacatca caaaattcat
181 tacaaaatgt cacttactgc tccatgctgg agaaagccat atccttctgg gacttgagtc
241 ggcacattta actacaggta ctgatctgtt ttgtgcttag ATGTTCCCCA TCATTGCCCA
301 GTATGGAGAT GTATTGGTGA GAAACTTGAG GCGGGAAAGCA GAGAAAGGCAG AGCCTGTAC
361 CTTGAAAGAG taagtaggag cacagccatg gggttcttag ctgtcatgag cccttcoagc
421 tgcctgccat ggagtgcaca gtgcactgt tgggttactc cagtgaccag aaaaaagcag
481 ggcagcgctg caactccaaa gagccaccta agagggagtg gtccccatga ggccggcaagt
541 cagcaaggaa aaaggccctt ctctcctgtg cacaggagcc aggatttact tatctgttaa
601 ctt

```

Seq ID: 136

>chzq\_ch-v-010\_132-133insGTC

```

1  attggacatg atagctagat ttgtttcagg aaaacatcct gctttccaag gatttagatg
61  aatgtttttt ttcactgggt actcaggtaa cacgtttca agaagccata gggaggttga
121 gggagggaaag tgcgtaagaa gggaggttga ggactgcact ttgttattac ttctgaettc
181 acgagtcact ttctgcccaaa gaaatctctc cttttgttc tagCACCGAC TAGATTCC
241 TCAGCTGATG ATTGACTCCC AGAATTCGAA AGAAACTGAG TCCCACAAAG gtaaccaagg
301 agtgcttctg agggctactg gcggggacac taagagggag ggccttgttc tgaaaatgtg
361 caggaagtat tccaggaaga tgagaatttt tgccacatag cagaacaaca cacattaga
421 tgttataaat ggtagctgga ggcactttcc agaagcccac agtatatgcc atgttccagg
481 ctgaaaggc aaccctaagc aaaccttagaaa tgcttggagg acagtcagt gtttgtggat
541 cacttatcatg agatcaaatg ccagttctca gcctccctcca gatccaccaa gtgagaacct
601 ctacttgaa atttatatca aacata

```

Seq ID: 137

>chzq\_ch-v-011\_364G>T

```

1  attggacatg atagctagat ttgtttcagg aaaacatcct gctttccaag gatttagatg
61  aatgtttttt ttcactgggt actcaggtaa cacgtttca agaagccata gggaggttga
121 gggagggaaag tcaagaaggg aggttggagga ctgcactttt gatttacttc tgacttcacg
181 agtcactttc tgccaaagaa atctctctt ttgcttcttag CACCGACTAG ATTCCTTCA
241 GCTGATGATT GACTCCCAGA ATTCGAAAGA AACTGAGTCC CACAAAGgta accaaggagt
301 gcttctgagg gctactggcg gggacactaa gagggagggc ctgttctga aaatgtgccg
361 gaatatttcc aggaagatgg gaatttttgc cacattgcg aacaacacac atttagatgt
421 tataaatgtt agctggaggc actttccaga agcccacagg tatagccatg ttccagggt
481 aaaggcaac cctaagccaa cctagaatgtc ttggaggaca gtcagttggtt tgtggatcac
541 ctacatgaga tcaaatgcca gttctcagcc tcctcccatg ccaccaatgtg agaacctcta
601 cttggaaatt tatatcaaac ata

```

Seq ID: 138

>chzp\_ch-v-012\_269G>A

```

1  agataaaagta cttttaggat cattcaaggc acacacccat aacactgagt atgtaagaca
61  gaaatgctt ctctgaaat tacagcatgtg ctggtctgg gatgccatga tgaggaggt
121 gtggcccaca atcatgtaga ctttggaaac acctggatta aatgatttt tcgtcatcct
181 ggccctgtat aagatacata tcagaatgaa aaccactccc agtgtgactt tgaattgctt
241 ttccatttt tcttcttggg attagagagc ttcacttta gctgtgatgt
301 tgttacgttg a cctgatttac cttaaatgtc ttcccttcc tttcagCTCT GTCTGATCTG
361 GAGCTCGCAG CCCAGTCATA ATCTTCATT TTTGCTGGCT ATGAAACCAC CAGCAGTGTT
421 CTTTCCTTCA CTTTATATGA ACTGGCCACT CACCCTGATG TCCAGCAGAA ACTGCAAAG
481 GAGATTGATG CAGTTTGCC CAATAAGgtg agggatgac cccttgagat gaagggaaga
541 ggtgaaggct tagaaaaat gcctcctcac cactccccag gagaattttt ataaaaagca
601 taatcactga ttccttcat gacataatgt aggaagcct tcgaggagaaaa aacaaaggga
661 gaaacataga gaacggttgc tactggcaga agcataagat ctttgtacaa tattgctggc
721 cctggttcac ctgtttactg ttatcacaat aatgcttaatg aaaaaaaaaaaaaaaa

```

11/21  
Figure 4 continued

781 aaaaaaaaaa aggagtgtgg cgagaagatg gccaaacagg aacagg

Seq ID: 139

>chzm\_ch-v-013\_167A>G

```

1 gtcctgggg tgaggatgtt cttgaatattc tcctacattc ataactcctc cacacatctc
61 agtaggtcac tgagcacatc aatggacatg ccagttatta aaatacttca cgaatactat
121 gatcatttac cagttatgtt tattctctgg agcttcaat acttcagt tactgcattgg
181 actcagttga gaggtaattc aaaatctcg attatccat tctgtttctt tccttccagG
241 CACCACTAC CTATGATGCC GTGGTACAGA TGGAGTACCT TGACATGGTG GTGAATGAAA
301 CACTCAGATT ATTCCCAGTT GCTATTAGAC TTGAGAGGAC TTGCAAGAAA GATGTTGAAA
361 TCAATGGGGT ATTCAATTCCC AAAGGGTCAA TGTTGGTGAT TCCAACTTAT GCTCTTCACC
421 ATGACCCAAA GTACTGGACA GAGCCTGAGG AGTTCCGCC TGAAAGgtac aagtctccag
481 ggaaatggag ctcaccctga cccaggctgg ttcaagcata ttctgcctct cttaatctac
541 atgacaatcg tgtgggtgtt caatcatttg ttgttaagt ttttatcac aaaaaagtga
601 taattatcaa actttacaaa ccacagacta gaaaaaacga aactacatcc atccacagtc
661 ccagcacaag acaaagataa tcaattatgt ccctgtggc attttctac gcctatata
721 atttttaaaa attagaatgg tatcactttt tatttgggtt gaattgctgc ttacttgatt
781 taacaggaaa ctatc

```

Seq ID: 140

>chzm\_ch-v-017\_248-249inst

```

1 gtcctgggg tgaggatgtt cttgaatattc tcctacattc ataactcctc cacacatctc
61 agtaggtcac tgagcacatc aatggacatg ccagttatta aaatacttca cgaatactat
121 gatcatttac cagttatgtt tattctctgg agcttcaat acttcaatag tactgcattgg
181 actcagttga gaggtaattc aaaatctcg attatccat tctgtttctt tccttccagG
241 CACCACTA CCTATGATGC CGTGGTACAG ATGGAGTACCT TGACATGGT GGTGAATGAA
301 ACACCTCAGAT TATTCCCAGT TGCTATTAGA CTTGAGAGGA CTTGCAAGAAA AGATGTTGAA
361 ATCAATGGGG TATTCAATTCC CAAAGGGTCA ATGGTGGTGA TTCCAACCTTA TGCTCTTCAC
421 CATGACCCAA AGTACTGGAC AGAGCCTGAG GAGTTCCGCC CTGAAAGgtac caagtctcca
481 ggaaatgga gctcaccctg acccaggctgg ttcaagcat attctgcctc tcttaatcta
541 catgacaatc gtgtgggtgtt acaatcattt gctttaagt ttttatca aaaaaagtga
601 ataattatca aactttacaaa accacagact agaaaaaacga aactacatc catccacagtc
661 cccagcacaac gacaaagata atcaattatgt ccctgtggc cattttctca cgcctatata
721 gatttttaaaa attagaatgg gtatcactttt tatttgggtt tgaattgctg cttacttgat
781 ttaacaggaaa actatc

```

Seq ID: 141

>peptide of ch-v-017\_248-249inst

1 ALSDLELAAQ SIIIFIFAGYE TTSSVLSFTL YELATHPDVQ QKLQKEIDAV LPNKAPPYLY.

Seq ID: 142

>chzm\_ch-v-014\_643C>T

```

1 gtcctgggg tgaggatgtt cttgaatattc tcctacattc ataactcctc cacacatctc
61 agtaggtcac tgagcacatc aatggacatg ccagttatta aaatacttca cgaatactat
121 gatcatttac cagttatgtt tattctctgg agcttcaat acttcaatag tactgcattgg
181 actcagttga gaggtaattc aaaatctcg attatccat tctgtttctt tccttccagG
241 CACCACTAC CTATGATGCC GTGGTACAGA TGGAGTACCT TGACATGGTG GTGAATGAAA
301 CACTCAGATT ATTCCCAGTT GCTATTAGAC TTGAGAGGAC TTGCAAGAAA GATGTTGAAA
361 TCAATGGGGT ATTCAATTCCC AAAGGGTCAA TGTTGGTGAT TCCAACTTAT GCTCTTCACC
421 ATGACCCAAA GTACTGGACA GAGCCTGAGG AGTTCCGCC TGAAAGgtac aagtctccag
481 ggaaatggag ctcaccctga cccaggctgg ttcaagcata ttctgcctct cttaatctac
541 atgacaatcg tgtgggtgtt caatcatttg ttgttaagt ttttatcac aaaaaagtga
601 taattatcaa actttacaaa ccacagacta gaaaaaacga aattacatcc atccacagtc
661 ccagcacaac acaaagataa tcaattatgt ccctgtggc attttctac gcctatata
721 atttttaaaa attagaatgg tatcactttt tatttgggtt gaattgctgc ttacttgatt
781 taacaggaaa ctatc

```

**12/21**  
**Figure 4 continued**

Seq ID: 143

&gt;chzn\_ch-v-015\_351T&gt;C

```

1 caggcctggc acagagttag tgctccataa atattttgtt aaacgatggaa tggtgagtg
61 tttaactatc cagtatttac ccagcttata gattaagtat gaagagttca agatacatgg
121 tgtaaagagt cgttttata tgcttgcaaa gcattttgt catatttttt ctactttgt
181 tccatctttt cttcttcac ttcatattt aattctccat atgcttgttt aactattgt
241 gATCCCCTTG AAATTAGACA CGCAAGGACT TCTTCACCA GAAAAACCCA TTGTTCTAAA
301 GGTGGATTCA AGAGATGGAA CCCTAAGTGG AGAATGAgtt attctaagga tttctacttt
361 ggtcttcaag aaagctgtgc cccagaacac cagagattt aacttagtca ataaaacctt
421 gaaataaaaga tgggcttaat ctaatgtact gcatgagtag ttggtgattt tgtacattca
481 ttgagctctc ccagagtctg ttagagtg ttagcattat gtatataaaa ggaggtgacc
541 aggttaagtga cagataggta gactcagttt ctctgcttcataggacta cctctaccca
601 cctctagttt gcat
```

Seq ID: 145

&gt;peptide of ch-v-001\_406C&gt;A

```
..1 PVAIRLERTC KKDVEINGVF IPKGSMVVIPI NYALHHDPKY WTEPEEFRPE RFSKKKDSID
```

Seq ID: 206

&gt;chyu\_ch-v-048\_206G&gt;A

```

1 cattttagtcc ttgtgagcac ttgtatgattt acctgccttc aatttttac tgacctaata
61 ttcttttta taatgaagta ttttaaacat ataaaacatt atggagagtgc gcataggaga
121 tacccacgta tgtaccaccc agcttaacgta atgctctact gtcatttcta accataatct
181 ctttaaagag ctctttgtc tttcatate tcttcctgt ttggaccaca ttaccctca
241 tcatatgaag ctttgggtgg ctcctgtgtg agactcttgc tttgtgtcac accctaata
301 actagaacct aagggtgtc tttgtgtcac aactagggtt atggattaca taacataatg
361 atcaaagtctt ggcttcctgg gtgtggctcc agctgcagaa tcgggcttagt gaagttat
421 cagctccgtt gtcacacac agAA
```

Seq ID: 207

&gt;chzk\_ch-v-037\_230T&gt;C

```

1 tggtcaccca ccatgtgtac agtaccctgc tagggtccag ggtcatgaaa gtaaataata
61 ccagactgtg cccttgagga actcacctct gctaagggaa acaggcacag aaacccacaa
121 gggtgttata gaggaaatag gacaatagga ctgtgtgagg gggataggag gcacccagag
181 gaggaaatgg ttacatctgt gtgaggaggt tggtaaggaa agactttaad agaaggggtc
241 tgtctggctg ggcttgcaag gatgtgttagg agtcatctag ggggcacaag tacactcag
301 gcagaggaa ttgcattgggaa aagatctgc agttgtggct tttggggatg gattcaagt
361 attctggaaat gaagacagcc atggaaacaa gggcaggta gaggatattt aagaggctc
421 atgccaatgg ctccacttca gtttctgata agaactcagg tttccgtggac tccctgataa
481 aactgattaa gtttttatg attccccata gaatatgaac tcaaaggagg taagcaaagg
541 ggtgtgtgcg attctttgtc actggctgca gctgcagccc cacctcccttc tccagcacat
601 aaacatttca gcagttgac ctaagactgc tttgtgtcac agggatgtc caggcagaca
661 gcccagcaaa caacagcaca cagctgaaaag taagactcag aggagacagt tgaagaaggc
721 aagtggcgat ggacttcattc ccaaatttgg cggtgaaac ctggcttctc ctggctgtca
781 gcctgggtgtc cctctatctg tttttttttt tccagattcc tccctctgt
```

13/21

**Figure 4 continued**

Seq ID: 208

```
>chzk_ch-v-038_506C>T
 1 tggtcaccca ccatgtgtac agtaccctgc tagggtccag ggtcatgaaa gtaaataata
 61 ccagactgtg cccttgagga actcacctct gctaaggaa acagggcacag aaacccacaa
121 ggggtttaga gagaaatag gacaatagga ctgtgtgagg gggataggag gcacccagag
181 gagaaatgg ttacatctgt gtgaggaggt tggtaaggaa agacttaat agaagggtc
241 tgtctggctg ggcttgcag gatgtgttagg agtcatctag gggcacaag tacactccag
301 gcagggaaa ttgcattgggaa aagatctgc agttgtggct tggggatg gattcaagt
361 attctggaat gaagacagcc atggaaacaa gggcagggtga gaggatattt aagaggctc
421 atgccaatgg ctccacttca gtttctgata agaactcagg ttccgtggac tccctgataa
481 aactgattaa gttgtttatg attcccata gaatatgaaac tcaaaggagg taagcaaagg
541 ggtgtgtgcg attctttgtc actggctgca gctgcagccc cacccttc tccagcacat
601 aaacatttca gcagcttgcac ctaagactgc tgcaggggc aggatgctc caggcagaca
661 gcccagcaaa caacagcaca cagctgaaag taagactcag aggagacagt tgaagaaggc
721 aagtggcgat ggacccatc ccaaatttgg cggtggaaac ctggcttc ctggctgtca
781 gcctgggtct cctctatctg tcaagtaactg tccagattcc tctcctctgt
```

Seq ID: 209

```
>chzk_ch-v-039_514T>C
 1 tggtcaccca ccatgtgtac agtaccctgc tagggtccag ggtcatgaaa gtaaataata
 61 ccagactgtg cccttgagga actcacctct gctaaggaa acagggcacag aaacccacaa
121 ggggtttaga gagaaatag gacaatagga ctgtgtgagg gggataggag gcacccagag
181 gagaaatgg ttacatctgt gtgaggaggt tggtaaggaa agacttaat agaagggtc
241 tgtctggctg ggcttgcag gatgtgttagg agtcatctag gggcacaag tacactccag
301 gcagggaaa ttgcattgggaa aagatctgc agttgtggct tggggatg gattcaagt
361 attctggaat gaagacagcc atggaaacaa gggcagggtga gaggatattt aagaggctc
421 atgccaatgg ctccacttca gtttctgata agaactcagg ttccgtggac tccctgataa
481 aactgattaa gttgtttatg attcccata gaacata tcaaaggagg taagcaaagg
541 ggtgtgtgcg attctttgtc actggctgca gctgcagccc cacccttc tccagcacat
601 aaacatttca gcagcttgcac ctaagactgc tgcaggggc aggatgctc caggcagaca
661 gcccagcaaa caacagcaca cagctgaaag taagactcag aggagacagt tgaagaaggc
721 aagtggcgat ggacccatc ccaaatttgg cggtggaaac ctggcttc ctggctgtca
781 gcctgggtct cctctatctg tcaagtaactg tccagattcc tctcctctgt
```

Seq ID: 210

```
>chzh_ch-v-051_455T>G
 1 caaaaattag ctaggtgtgg tggtatgcac ctataatctc agctacccag gaggctgagg
 61 caggagaatc acttgaacct ggagggcagag gttgcagtga gccgagacgc accattacac
121 tccagcctgg gtgacagagt gagattccat ctaaaaaaaaaaaaaaa attatgcctt
181 tttgaagcac atacattttta taacatacaa ctgaatccct tattatatta ttatgtttga
241 ttaatgttt tcaaaccatc tcccctgata ttctggag atggaaaca tgttttctta
301 cacctctgc attccattct caactcccaa ctgtctact gcaatgaaca cttataaaga
361 aacagtcaat tggtaatttgc attgggcaac aggctaaaca cactcattcc ttgtctgtc
421 ccactcttt ctttactttc ctttctctgag taactatcc taaaagtattt aggtgggtgg
481 cagccagatg gtggccacac attaaggtag aaaagagagt gtcatgtatgg ttccaagtca
541 gagaccttagt agggtgagga tcaagtaggt gttcacgtgg agaaacagcc cggcctgtgt
601 gtgggaggtcc aagcaagcag agaaaatgtc gacacagagg ggtggctga aaaagcagcc
661 agagcctaaa cagggcatgg agaacaatatt tagggcatga ggtgaggagg gcatccatga
721 gtgggaaaggc atgggtgagg ttcaactaca taaaagggtat tgatgaaata agtaaataaa
781 gtatactgga agccaggtgt gtcactttt cagaaaagag tcatggattc agaaagg
```

14/21

## Figure 4 continued

Seq ID: 211

&gt;chzh\_ch-v-052\_577G&gt;A

```

1 caaaaattag ctaggtgtgg tggatgcac ctataatctc agctaccagg gaggctgagg
61 caggagaac acttgaacct ggaggcagag gttcaagtga gccgagacgc accattacac
121 tccagcctgg gtgacagagt gagattccat ctaaaaaaaaaaaa attatgcctt
181 tttaaagcac atacatTTTA taacatacaa ctgaatccc tattatatta tttagtttga
241 tttaatgtt tcaaaaccat tcccctgata ttctggag atggaaaca tggtttctta
301 cacctcttc attccatttc caactccaa ctgtctact gcaatgaaca cttataaga
361 aacagtcaat tggtaatttggatggcaac aggctaaaca cactcattcc ttgtctgttc
421 ccacttctt ctttacttcc ctttcctgag taacttacc taaagtcatgg atgtgggtgg
481 cagccagatg gtggccacac attaaaggtag aaaagagagt gtcatgatgg ttccaagtca
541 gagacctagt agggtgagga tcaagtaggt gttcaactgtt agaaacagcc cggcctgtgt
601 gtgggagtcc aagcaagcag agaaaaatgtc gaçacagagg ggtggcctga aaaagcagcc
661 agagcctaaa cagggcatgg agaacatatt tagggcatga ggtgaggagg gcatccatga
721 gtgggaagggg atgggtgagg ttcaactaca taaagggtt tgatgaaata agtaaataaa
781 gtatactgga agccagggtgt gtcactttt cagaaaagag tcatggattc agaaagg
```

Seq ID: 212

&gt;chzw\_ch-v-053\_163G&gt;A

```

1 catagacaag ggtgagtcct tcagttactt gaaaaatttcc aagagtgact taaaattccc
61 cacttcaaat atattctctg ttttcttgc tttcccttaa gacatctctg aatagctcc
121 ttcaactgcc agtgaagat agcaggcctg atttcattgg accaactgt tttcagcccc
181 aatttagaggt agggtttatt ctatTTTTAA taataatcaa ctgttatTTT gtttcctctc
241 ccagGGTCTC TGGAAATTG ACACAGAGTG CTATAAAAAG TATGGAAAAA TGTGGGGGtg
301 agtattctga aaacotccat tggatagacc tgctactgtg aggaggttac cccactgcag
361 gatagtcttcccccaggtct tcattggatg aagctttgtt caacctaaat acaaacagag
421 agagggttcc ttgaaagaaga ggataattac ttgggagtag aatattgcaa tgggaatctg
481 cttgcgtta taaactatgt gcaatttccatggatg aagctttgtt aagacaaaga tgctccatag
541 aaaatatgag aagaatctca taactgtttt gagataatta ttgttagcta caaagatcaa
601 ta
```

Seq ID: 213

&gt;chzw\_ch-v-054\_444T&gt;A

```

1 catagacaag ggtgagtcct tcagttactt gaaaaatttcc aagagtgact taaaattccc
61 cacttcaaat atattctctg ttttcttgc tttcccttaa gacatctctg aatagctcc
121 ttcaactgcc agtgaagat agcaggcctg atttcattgg accaactgt tttcagcccc
181 aatttagaggt agggtttatt ctatTTTTAA taataatcaa ctgttatTTT gtttcctctc
241 ccagGGTCTC TGGAAATTG ACACAGAGTG CTATAAAAAG TATGGAAAAA TGTGGGGGtg
301 agtattctga aaacotccat tggatagacc tgctactgtg aggaggttac cccactgcag
361 gatagtcttcccccaggtct tcattggatg aagctttgtt caacctaaat acaaacagag
421 agagggttcc ttgaaagaaga ggaattac ttgggagtag aatattgcaa tgggaatctg
481 cttgcgtta taaactatgt gcaatttccatggatg aagctttgtt aagacaaaga tgctccatag
541 aaaatatgag aagaatctca taactgtttt gagataatta ttgttagcta caaagatcaa
601 ta
```

Seq ID: 214

&gt;chzt\_ch-v-043\_294T&gt;C

```

1 agcgaaaaac tcaaggaggt atgaaaataa gatgagtctt aattagaaat gtaaagaatg
61 aatctgggaa caggtagaaa gtaagatcac agtccgtttc caaggggttag tccactgagt
121 tgcgttcc taaaatgtt ctttatctt tatgtacaga aaagacatca caaaattcat
181 tacaaaaatgt cacttactgc tccatgctgg agaaaaggccat atccctctgg gacttgagtc
241 tgcacatttactacaggta ctgtatgtt ttgtgttttag ATGTTCCCCCA TCATGCCCA
301 GTATGGAGAT GTATTGGTGA GAAACTTGAG GCGGGAAAGCA GAGAAAGGCA AGCCTGTCAC
361 CTTGAAAGAg taagttaggag cacagccatg gggttctgag ctgtcatgag cccttccagc
421 tgcctgcatg ggagtcgaca gtcgcactgt tgggttactc cagtgaccag aaaaaagcag
481 ggcagcgtcaactccaa gagccaccta agaggaggtg gctccatga ggcggcaagt
541 cagcaaggga aaaggccctt ctctctgtt cacaggagcc aggtttact tatctgtt
```

15/21

## Figure 4 continued

601 ctt

Seq ID: 215

&gt;peptide of ch-v-043\_294T&gt;C

1 SLAEDEEWKR IRSLLSPTFT SGKLKEMFPI **T**AQYGDVLVR NLRREAEKGK PVTLKDIFGA

Seq ID: 216

&gt;chzt\_ch-v-055\_444G&gt;A

1 agcggaaaac tcaaggaggt atgaaaataa gatgagtctt aattagaaaat gtaaaaatg  
 61 aatctgggaa caggtagaaa gtaagatcac agtccgttc caaggggtag tccactgagt  
 121 tgcagcttcc taaaaatggt cttttatctt tatgtacaga aaagacatca caaaattcat  
 181 tacaaaatgt cacttactgc tccatgctgg agaaaaggcat atccttctgg gacttgagtc  
 241 tgcacattta actacaggta ctgatctgtt ttgtgcttag ATGTTCCCCA TCATTGCCCA  
 301 GTATGGAGAT GTATTGGTGA GAAACTTGAG GCGGGAAAGCA GAGAAAGGCA AGCCTGTCAC  
 361 CTTGAAAGAg taagtaggag cacagccatg gggttctgag ctgtcatgag cccttccagc  
 421 tgccctgcat ggagtcgaca gtcactgt tgggttactc cagtgaccag aaaaaagcag  
 481 ggcagcgctg caactccaaa gagccaccta agagggagtg gctccatga ggccgcaagt  
 541 cagcaaggaa aaaggccctt ctctcctgtg cacaggagcc aggatttact tatctgttaa  
 601 ctt

Seq ID: 217

&gt;chzs\_ch-v-050\_437G&gt;A

1 cagagcttac atatcttata tcatccacac tcaacacatg ctactgttagt tgtctgataa  
 61 tgggtctctg tcttcctatg actgggctcc ttgacccataggtgagtct aactcagctt  
 121 ggtgtctcca tcaccccccag catagggcca gctccatcac tggcaccaga taaccacctt  
 181 ctgagggagt agatggaaga tgattcagca gatagttctg aaagtctgtg gctctttag  
 241 tgtcttgact ggataatgtgg gtttcttgc gcatgtatag tggaggacg gtaagaggtg  
 301 ctgatTTaa tttccatat ctttctccac tcagCATCTT TGGGGCCTAC AGCATGGATG  
 361 TGATTACTGG CACATCATTT GGAGTGAACA TCGACTCTCT CAACAATCCA CAAGACCCT  
 421 TTGTGGAGAG CACTAAAAG TTCCTAAAAT TTGGTTCTT AGATCCATTA TTTCTCTCAA  
 481 TAAgtatgtg ggctattatt tctttctctc tttttaaaaa taactgctt cttgacatat  
 541 aattcacata tcgtataatt catccactta aaaggtacaa ttccattgtt tttaagataa  
 601 tcaaaaatgt gatgaccat tactattgt aactaaaatg tttttgtcaa tctagagccc  
 661 tcacacac tt tagctgtcaa cacccacca caaaccac tggccctaagc atccaataat  
 721 caactttctg cctctataga ttgccttatt ctggacactt catagaaata atatcatt

Seq ID: 218

&gt;chzs\_ch-v-056\_467A&gt;G

1 cagagcttac atatcttata tcatccacac tcaacacatg ctactgttagt tgtctgataa  
 61 tgggtctctg tcttcctatg actgggctcc ttgacccataggtgagtct aactcagctt  
 121 ggtgtctcca tcaccccccag catagggcca gctccatcac tggcaccaga taaccacctt  
 181 ctgagggagt agatggaaga tgattcagca gatagttctg aaagtctgtg gctctttag  
 241 tgtcttgact ggataatgtgg gtttcttgc gcatgtatag tggaggacg gtaagaggtg  
 301 ctgatTTaa tttccatat ctttctccac tcagCATCTT TGGGGCCTAC AGCATGGATG  
 361 TGATTACTGG CACATCATTT GGAGTGAACA TCGACTCTCT CAACAATCCA CAAGACCCT  
 421 TTGTGGAGAG CACTAAAAG TTCCTAAAAT TTGGTTCTT AGATCCGTTA TTTCTCTCAA  
 481 TAAgtatgtg ggctattatt tctttctctc tttttaaaaa taactgctt cttgacatat  
 541 aattcacata tcgtataatt catccactta aaaggtacaa ttccattgtt tttaagataa  
 601 tcaaaaatgt gatgaccat tactattgt aactaaaatg tttttgtcaa tctagagccc  
 661 tcacacac tt tagctgtcaa cacccacca caaaccac tggccctaagc atccaataat  
 721 caactttctg cctctataga ttgccttatt ctggacactt catagaaata atatcatt

16/21

### Figure 4 continued

Seq ID: 219

>chzs ch-v-057 583C>T

```
1 cagagcttac atatcttata tcatccacac tcaacacatg ctactgttagt tgtctgataa  
61 tgggtctctg ttccctatg actgggctcc ttgacctcag aggtagtct aactcagctt  
121 ggtgtctcca tcacccccag catagggcca gctccatcac tggcaccaga taaccacctt  
181 ctgagggagt agatggaga tgattcagca gatagttctg aaagtctgtg gctctttatg  
241 tgcttgcact ggatatgtgg gtttcttgc gcatgtatag tggaggacg gtaagagggt  
301 ctgattttaa ttttccatat ctttcac tcagCATCTT TGGGGCCTAC AGCATGGATG  
361 TGATTACTGG CACATCATTT GGAGTGAAACA TCGACTCTCT CAACAATCCA CAAGACCCCC  
421 TTGTGGAGAG CACTAAGAAG TTCCTAAAAT TTGGTTTCTT AGATCCATTA TTTCTCTCAA  
481 TAAgtatgtg ggcttattatt tctttctctc tttttaaaaa taactgctt cttgacatata  
541 aattcacata tcgtataatt catccactta aaaggtacaa tttcattgtt ttaagataaa  
601 tcaaaaatat gtatgaccat tactattgt aactaaaatg ttttgcataa tctagagccc  
661 tcacacactt tagctgtcaa caccacca caaacccac tgccctaagc atccaataat  
721 caactttctg cctctataga ttgcctatt ctggacactt catagaaata atatcatt
```

Seq ID: 220

>chzs ch-v=058 650A>G

```
1 cagacttac atatctata tcatccacac tcaacacatg ctactgttagt tgtctgataa  
61 tgggtctctg tcttcctatg actgggctcc ttgacctcag agggtgagtct aactcagtt  
121 ggtgtctcca tcacccccag catagggcca gctccatocac tggcaccaga taaccacctt  
181 ctgagggagt agatggaga tgatttcgca gatagtttgc aaagtctgtg gctctttatg  
241 tgccttgact ggatatgtgg gtttcttgc gcatgtatag tggaaaggacg gtaagaggtg  
301 ctgattttaa tttccatat ctttccac tcagCATCTT. TGGGGCCTAC AGCATGGATG  
361 TGATTACTGG CACATCATTT GGAGTGAAACA TCGACTCTCT CAACAATCCA CAAGACCCCT  
421 TTGTGGAGAG CACTAAGAAC TTCCTAAAAT TTGGTTTCTT AGATCCATT A TTTCTCTCAA  
481 TAAgtatgtg ggcttatttt tctttctctc tttttaaaaaa taactgcttt cttgacatata  
541 aattcacata tcgtataatt catccactta aaaggtacaa ttccattgtt tttaagataaa  
601 tcaaaaatat gtatgaccat tactattgtt aactaaaatg tttttgtcag tctagagccc  
661 tcacacactt tagctgtcaa caccacca caaacccac tgccctaagc atccaataat  
721 caactttctg cctctataga tttgcctatt ctggacactt catagaaaata atatcatt
```

Seq ID: 221

>chzr ch-v-059 205T>C

1 agaagggtgcc attgatctca ctgctgttagt ggtgtttcct atgtatagac ctggcccttg  
61 tcagtgcgg gcctgaaaga agggcaaaca tgat~~aaa~~agg aatgggttcc agttgagaat  
121 catatgtttc ttattcttat tactggtaga gaaaattata attgctccag gtaaaatttg  
181 catttcaat gatttccccc tgg~~t~~ttttt gtttttccca cagTACTCTT TCCATTCCCT  
241 ACCCCAGTTT TTGAAGCATT AAATGTCTCT CTGTTTCCAA AAGATACCAT AAATTTTTTA  
301 AGTAAATCTG TAAACAGAAT GAAGAAAAGT CGCCTCAACG ACAAACAAAA Gttaaaatct  
361 gatgggtggtt aaatgacgat gtttaggttt tgataaaattt agatttata cacatgatag  
421 agcatgttac tggat~~tttta~~ aaaataaaga cagagaactt atgttttagaa caagagaagc  
481 catttggtag aaataaagaa ggagattggg gaaggagatg agaatgagtc agagagatag  
541 catttaaaac ttgaaatcg gcacaacaat tagtatgtca tgatataaac agtattgaga  
601 taaaattttt ccacttctct tccctttaat aaattgtcaa aggataaagt ttccctgtttg  
661 aaaatatatt

17/21

**Figure 4 continued**

Seq ID: 222

>chzr\_ch-v-060\_496A>C

1 agaagggtcc attgatctca ctgctgttagt ggttccct atgtatagac ctgccttgc  
61 tcagtcggc gcctgaaaaga aggccaaaca tgataaaagg aatgggttcc agttgagaat  
121 cattatgttc ttattcttat tactggtaga gaaaattata attgctccag gtaaagtttg  
181 catttcaat gatttcctt tgtttgggtt gttttccca cagTACTCTT TCCATTCCCTT  
241 ACCCCAGTTT TTGAAGCATT AAATGTCTCT CTGTTTCCAA AAGATACCAC AAATTTTTTA  
301 AGTAAATCTG TAAACAGAAT GAAGAAAAGT CGCCTCAACG ACAAAACAAAA Gttaaatct  
361 gatggtggtt aaatgacgat gtttaggttt tgataaaattt agattttata cacatgatag  
421 agcatgtatc tgtatTTTA aaaataaaga cagagaactt atgTTTtagaa caagagaagc  
481 cattggtag aaataagaa ggagattggg gaaggagatg agaatgagtc agagagatag  
541 cattaaaaac ttgaaatcag gcacaacaat tagtatgtca tgatataaac agtattgaga  
601 taaaatttta ccacttctct tcccttaat aaattgtcaa aggataaagt ttccctgtttg  
661 aaaatatatt

Seq ID: 223

>chzp\_ch-v-062\_173G>A

1	agataaaagt	cttttaggat	cattcaaggc	acacacccat	aacactgagt	atgtaaagaca
61	gaaatgcct	ctctggaaat	tacagcagt	ctgggtctgg	<u>gatgccatga</u>	<u>tgaggaggtgt</u>
121	gtggcccaca	atcatgtaga	ccttggaaa	acctggatta	aatatgttt	gd <u>atcatct</u>
181	ggccctgtat	aagatacata	tcagaatgaa	aaccactccc	agtgtgactt	tgaattgctt
241	ttccatffff	tcttcttggg	attagagagc	ttcactttaga	tttcatttaa	gctgtgtatgt
301	tgtacgttga	cctgattttac	ctaaaatgtc	tttcctctcc	tttcagCTCT	GTCTGATCTG
361	GAGCTCGCAG	CCCAGTCAAT	AATCTTCATT	TTTGCTGGCT.	ATGAAACAC	CAGCAGTGTG
421	CTTCCTTCA	CTTTATATGA	ACTGCCACT	CACCCGTATG	TCCAGCAGAA	ACTGAAAAG
481	GAGATTGATG	CAGTTTGCC	CAATAAGgtg	aggggatgac	ccctggagat	gaaggaaaga
541	ggtaaggct	tagaaaaat	gccttcac	cactccccag	gagaattttt	ataaaaagca
601	taatcactga	ttccttcaact	gacataatgt	aggaagcctc	tgaggagaaaa	aacaaaggga
661	gaaaataga	gaacggttgc	tactggcaga	agcataagat	cttgtacaa	<u>tattgtctggc</u>
721	<u>cctggttcac</u>	ctgtttactg	ttatcacaat	aatgctaagt	aaaaaaaaaa	aaaaaaaaaa
781	aaaaaaaaaa	aggagtgtgg	cgagaagatg	gccaaacagg	aacagc	

Seq ID: 224

>chzp ch-v-063. 312C>T

1 agataaaagta ctttttagat cattcaaggc acacacccat aacactgagt atgtaagaca  
61 gaaatgcctc ctctggaaat tacagcagt~~f~~ ctgggtctgg gatgccatga tgaggaggatgt  
121 gtggcccaca atcatgtaga ccttggaaa acctggatta aaatgat~~ttt~~ gcgtcatct  
181 ggcctgtat aagatacata tcagaatgaa aaccactccc agtgtgactt tgaattgtct  
241 ttccat~~ttttt~~ tcttcttggg attagagagc ttca~~ctt~~ tttcatctaa gctgtatgt  
301 tgtacgttga ct~~t~~gattac ctaaaatgtc ttcc~~c~~tc~~t~~cc tttcagCTCT GTCTGATCTG  
361 GAGCTCGCAG CCCAGTCAAT AATCTTCATT TTTGCTGGCT ATGAAACCAC CACCGAGTGT  
421 CTTCCCTTCA CT~~T~~TATATGA ACTGCCACT CACCC~~T~~GATG TCCAGCAGAA ACTGAAAAG  
481 GAGATTGATG CAGTTTGCC CAATAAGgtg aggggatgac ccctggagat gaaggaaaga  
541 ggtgaaggct tagaaaaat gccttc~~c~~cac cactccccag gagaat~~ttt~~ ataaaaaagca  
601 taatcactga ttcc~~t~~act gacataatgt aggaagcctc tgaggagaaa aacaaaggga  
661 gaaacataga gaacggttgc tactggcaga agcataagat ctttgataaa tattgtctggc  
721 cctggttcac ctgtttactg ttatcacaat aatgctaagt aaaaaaaaaa aaaaaaaaaa  
781 aaaaaaaaaa aggagtgtgg cgagaagatg gccaaacagg aacagc

18/21

## Figure 4 continued

Seq ID: 225

&gt;chzn\_ch-v-044\_239T&gt;C

```

1 caggcctggc acagagtcag tgctccataa atatttgtt aaacgatgga tggtgagtgc
61 tttactatac cagttttac ccagttata gattaagttt gaagagtca agatacatgg
121 tgttaagagt cgttttata tgcttgc当地 gcattttgtt catatttt ctactttgct
181 tccatctttt cttctttcac ttcatattt aattctccat atgcttggg aactattgca
241 gATCCCCTTG AAATTAGACA CGCAAGGACT TCTTCACCA GAAAAACCCA TTGTTCTAAA
301 GGTGGATTCA AGAGATGGAA CCCTAAGTGG AGAATGAgtt attctaagga tttctacttt
361 ggtcttcaag aaagctgtgc cccagaacac cagagatttc aacttagtca ataaaacctt
421 gaaataaaaga tgggtttaat ctaatgtact gcatgagtag ttggtgattt tgtacatca
481 tttagctcc ccagagtctg tgttaggtgt tgcattat gtatataaa ggaggtgacc
541 aggttaagtg aagatggta gactcagtt ctctgcttcatatggacta cctctaccca
601 cctctatgtt gatta

```

Seq ID: 226

&gt;chzy\_ch-v-066\_380T&gt;C

```

1 gattaagctt ttcatgattc ctcatagaac atgaactcaa aagaggtcag caaagggttg
61 tttgcgatc tttgttattt gctcgacta tagccctgcc tccttctcca gcacataaat
121 cttccagcag cttggctgaa gactgctgt cagggcaggg aagctccagg caaacagccc
181 agcaaacagc agcactcagc taaaaggaag actcacaagaa cacagttgaa gaaggaaagt
241 ggcgATGGAC CTCATCCCAA ATTTGGCGGT GGAAACCTGG CTTCTCCTGG CTGTCAGCCT
301 GGTGCTCCTC TATCTgtgag taactgtcca aactcccttc tttgtttctt tggacttggg
361 gtgctaatcg ggcccccttc cccttatctg tttgaagat caaaaagagat gttcaaggag
421 aagtagctga agtgttggac gctacaaacg catagaagtt attattatct tatgcagatc
481 tatgaatgaa taaataagca tttctcccat ccaccttcta attttggtga cttaggagggt
541 tttagggacag catttggtag tggaaatgtt tagatctgac gaagactaat
601 caatgaaaac atggcagcgg caga

```

Seq ID: 227

&gt;chzy\_ch-v-067\_474G&gt;A

```

1 gattaagctt ttcatgattc ctcatagaac atgaactcaa aagaggtcag caaagggttg
61 tttgcgatc tttgttattt gctcgacta tagccctgcc tccttctcca gcacataaat
121 cttccagcag cttggctgaa gactgctgt cagggcaggg aagctccagg caaacagccc
181 agcaaacagc agcactcagc taaaaggaag actcacaagaa cacagttgaa gaaggaaagt
241 ggcgATGGAC CTCATCCCAA ATTTGGCGGT GGAAACCTGG CTTCTCCTGG CTGTCAGCCT
301 GGTGCTCCTC TATCTgtgag taactgtcca aactcccttc tttgtttctt tggacttggg
361 gtgctaatcg ggcccccttc cccttatctg tttgaagat caaaaagagat gttcaaggag
421 aagtagctga agtgttggac gctacaaacg catagaagtt attattatct tatacatc
481 tatgaatgaa taaataagca tttctcccat ccaccttcta attttggtga cttaggagggt
541 tttagggacag catttggtag tggaaatgtt tagatctgac gaagactaat
601 caatgaaaac atggcagcgg caga

```

Seq ID: 228

&gt;chzu\_ch-v-068\_359G&gt;A

```

1 gaacagtcc ctaccacgtg gaggcatttc aattaaaaagg agactgagat atagaggcag
61 gagaccacac cagatggctg ggtctccccc ctcccacccc ccccccacat acactcagaa
121 gaggcttagc atctaggatc tccattgac atcttgcata tggcttgc当地 taatatcata
181 tacagtcata aatattttt taaataagga tgccttca atatatttt tgcaaccatg
241 aagatcacca caactaatgt gaaaaaaaat gtttctgtt aactctagTC TTTAGGCCA
301 GTGGGATTAA TGAAAAGTGC CATCTCTTAA GCTGAGGATG AAGAATGGAA GAGAATACAG
361 TCATTGCTGT CTCCAACCTT CACCAGCGGA AAACTCAAGG AGgtatgaaa ataagatgag
421 tcttaattag aatgtaaag aatgaatctg gggacaggtt gaaaatgttca tcacagtccg
481 tttccaagggtt gtagtccact gagttcgac ttcctaaaaa tggctttta tctttatgtt
541 cagaaaagac atcacaat tcaattacaa atgtcactt ctgctccatg ctggagaaag
601 ccatatccctt ctggactt ggtctgtt

```

19/21

## Figure 4 continued

Seq ID: 229

&gt;peptide of ch-v-068\_359G&gt;A

1 SVFTNRRSLG PVGFMKSAIS LAEDEEWKRI QSLLSPFTS GKLKEMFPII AQYGDVLVRN

Seq ID: 230

&gt;chzu\_ch-v-047\_404T&gt;C

1 gaacagttcc ctaccacgtg gaggcattgc aattaaaagg agactgagat atagaggcag  
 61 gagaccacac cagatggctg ggtctccccca ctcccacccc cgccccacat acactcagaa  
 121 gaggcttaggc atctaggatc tccattgagc atcttgaata tggcttgcca taatatcata  
 181 tacagtcaat aaatatttgt taaataagga tgcctttca atatatttg tgcaaccatg  
 241 aagatcacca caactaatgt gaaaaaaat gtttctgtt aactctagTC TTTAGGCCA  
 301 GTGGGATTAA TGAAAAGTGC CATCTCTTA GCTGAGGATG AAGAATGGAA GAGAATACGG  
 361 TCATTGCTGT CTCCAACCTT CACCAGCGGA AAACCTCAAGG AGgttatgaaa ataagatgag  
 421 tcttaattag aaatgtaaag aatgaatctg gggacaggtt gaaagtaaga tcacagtccg  
 481 ttccaagggtt gtagtccact gatgttcgagc tgcctaaaaa tggctttta tctttatgtt  
 541 cagaaaagac atcacaat tcattacaaa atgtcactt ctgcctcatg ctggagaaag  
 601 ccatatcctt ctggacttg agtctg

Seq ID: 231

&gt;chzu\_ch-v-069\_480G&gt;A

1 gaacagttcc ctaccacgtg gaggcattgc aattaaaagg agactgagat atagaggcag  
 61 gagaccacac cagatggctg ggtctccccca ctcccacccc cgccccacat acactcagaa  
 121 gaggcttaggc atctaggatc tccattgagc atcttgaata tggcttgcca taatatcata  
 181 tacagtcaat aaatatttgt taaataagga tgcctttca atatatttg tgcaaccatg  
 241 aagatcacca caactaatgt gaaaaaaat gtttctgtt aactctagTC TTTAGGCCA  
 301 GTGGGATTAA TGAAAAGTGC CATCTCTTA GCTGAGGATG AAGAATGGAA GAGAATACGG  
 361 TCATTGCTGT CTCCAACCTT CACCAGCGGA AAACCTCAAGG AGgttatgaaa ataagatgag  
 421 tcttaattag aaatgtaaag aatgaatctg gggacaggtt gaaagtaaga tcacagtccg  
 481 ttccaagggtt gtagtccact gatgttcgagc tgcctaaaaa tggctttta tctttatgtt  
 541 cagaaaagac atcacaat tcattacaaa atgtcactt ctgcctcatg ctggagaaag  
 601 ccatatcctt ctggacttg agtctg

Seq ID: 232

&gt;chzq\_ch-v-061\_194C&gt;G ...

1 attggacatg atagcttagat ttgttcagg aaaacatcct gcttccaag gatttagatg  
 61 aatgttttg ttcaactgggt actcaggtaa cacgtttca agaagccata gggaggttga  
 121 gggagggaaag tcaagaaggg agttgagga ctgcactttt gatttacttc tgacttcacg  
 181 agtcaacttc tgc~~aa~~agaa atctctcctt ttgctcttag CACCGACTAG ATTCCTTCA  
 241 GCTGATGATT GACTCCCAGA ATTCGAAAGA AACTGAGTCC CACAAAGgtt accaaggagt.  
 301 gcttctgagg gctactggcg gggacactaa gagggaggc ctgttctga aaatgtgcag  
 361 gaagtattcc aggaagatga gaattttgc cacatagcag aacaacacac atttagatgt  
 421 tataaatgggt agtggaggc actttccaga agcccacagg tatagccatg ttccaggctg  
 481 aaaggcaac cctaagcaaa cctagaatgc ttggaggaca gtca~~gt~~ggtt tggatcac  
 541 ctacatgaga tcaa~~at~~gcca gttctcagcc tcctccagat ccaccaagtg agaacctcta  
 601 cttggaaatt tataatcaa~~a~~ ata

Seq ID: 233

&gt;chzn\_ch-v-045\_291T&gt;C

1 caggcctggc acagagtcag tgctccataa atattttgtt aacgatgga tggtagtgc  
 61 ttactatc cagtattac ccagcttata gattaagtat gaagagttca agatacatgg  
 121 tgttaagagt cg~~tttttata~~ tgc~~tttgc~~aaa gcattttgt catat~~ttttt~~ ctactttgt  
 181 tccatctttt cttcttac ttcattttt aattctccat atgttgc~~ttt~~ aactattgtt  
 241 gATCCCCTTG AAATTAGACA CGCAAGGACT TCTTCAACCA GAAAAACCCA TGTTCTAAA  
 301 GGTGGATTCA AGAGATGGAA CCCTAAGTGG AGAATGAg~~t~~ttt attctaa~~g~~gg aacttacttt  
 361 ggtcttcaag aaagctgtgc cccagaacac cagagattc aacttagtca ataaaacctt

20/21

## Figure 4 continued

421 gaaataaaaga tgggcttaat ctaatgtact gcatgagtag ttggtgattt tgtacattca  
 481 tttagctc ccagagtctg tgttagtgt tgtgcattat gtatataaaa ggaggtgacc  
 541 aggttaagtga cagataggta gactcagtt ctctgcttct cataggacta cctctaccca  
 601 cctctagtta gcatta

Seq ID: 234

&gt;peptide of ch-v-045\_291T&gt;C

1 MRFALMNMKL ALIRVLQNFS FKPCKETQIP LKLDTQGLLQ PEKP~~T~~VLKVD SRDGTLSE.

Seq ID: 235

&gt;chzk\_ch-v-065\_563T&gt;C

1 tggtcaccca ccatgtgtac agtaccctgc tagggtccag ggtcatgaaa gtaataataa  
 61 ccagactgt cccttggagga actcaccct gctaagggaa acaggcacag aaacccacaa  
 121 ggggtgtaga gaggaaatag gacaatagga ctgtgtgagg gggataggag gcacccagag  
 181 gaggaaatgg ttacatctgt gtgaggaggt tggtaaggaa agactttaat agaaggggtc  
 241 tgtctggctg ggcttgcaag gatgtgtagg agtcatctag gggcacaag tacactccag  
 301 gcagagggaa ttgcatgggt aaagatctgc agttgtggct tgggggatg gattcagt  
 361 attctggaat gaagacagcc atggaaacaa gggcagggtga gaggatattt aaggggcttc  
 421 atgccaatgg ctccacttca gtttctgata agaactcagg ttccgggac tccctgataa  
 481 aactgattaa gttgttattg attccccata gaatatgaac tcaaaggaggg taagcaaagg  
 541 ggtgtgtgcg attctttgct acccggctgcg gctgcagccc cacctcccttc tccagcacat  
 601 aaacatttca gcagcttgcac ctaagactgc tgtgcagggc agggatgttc caggcagaca  
 661 gcccagcaa caacagcaca cagctgaaag taagactcag aggagacaggt tgaagaagggc  
 721 aagtggcgat ggaccttcatcc ccaaatttgg cggtggaaac ctggcttctc tggctgtca  
 781 gcctgggtgct cctcattctg tcagtactg tccagatcc ttcctctgtgt

Seq ID: 236

&gt;chz1\_ch-v-040\_206C&gt;T

1 ggagtgacct gatttccag gtgtgtctg tcacccctttt ctttgactag gaaaggggaaac  
 61 tccttgaccc cttgcgctttc tcaugtgagg caatgcctcg cctcgtttcg gcttgtgcac  
 121 agacacgtgc accccactgttc ctgcacccac tgtctggac tccctatgtga gatgaacccg  
 181 gtacctcaga tgaaaatgc aaaatcc gtcttcgtg tcactcacaac tgggagctgt  
 241 agacccggac tgttcctatt cggccatttt ggctccaccc cccgagttt ggcttttaat  
 301 tgaaaatgtta ttgatgtggg aaggagataa tgccatgtca ttatgaggcac atattagagg  
 361 gtctgagacca atgcatgtga taaaaggtca cctaagggaag aaaaaaaagaa caagggaaga  
 421 cactggaaag aacgtgtactc tgggagtcc tggccacca aagtctggag aaaaggtta  
 481 accacaaggc tccacgcca tgttcactga aagacctgcac actaggtactc ttatggatc  
 541 ctttggtagga caagccttga aaagttttaac aatagcaaa tgtggacgttg ttcagaccaa  
 601 atgatgttcac gtgtgttatttt gtgtgtgtgt gtcagtgtgt gtgtttaaaa atcatgacaa  
 661 ataaagcgac ctgtgaagag gggatccca tgctcggttg cctgataacca acactatcac  
 721 aaacgctttg cgaaacccca agtttgcaca aaggcaatcc caacctttaca caaaaa

21/21

## Figure 5

&gt;CYP3A5 insert in pKK233-2 (Seq ID 144)

```

1 ccATGgctct gttattagca gtttttCTGG TGCTCCTCTA TCTATATGGG ACCCGTACAC
61 ATGGACTTT TAAGAGACTG GGAATTCCAG GGCCCACACC TCTGCCTTTG TTGGGAAATG
121 TTTTGTCCTA TCGTCAGGGT CTCTGGAAAT TTGACACAGA GTGCTATAAA AAGTATGGAA
181 AAATGTGGGG AACGTATGAA GGTCAACTCC CTGTGCTGGC CATCACAGAT CCCGACGTGA
241 TCAGAACAGT GCTAGTGAAA GAATGTTATT CTGTCTTCAC AAATCGAAGG TCTTTAGGCC
301 CAGTGGGATT TATGAAAAGT GCCATCTCT TAGCTGAGGA TGAAGAATGG AAGAGAATAC
361 GGTCAATTGCT GTCTCCAACC TTCACCAGCG GAAAACCTAA GGAGATGTTG CCCATCATTG
421 CCCAGTATGG AGATGTATTG GTGAGAAACT TGAGGGGGG AGCAGAGAAA GGCAAGCCTG
481 TCACCTTGAA AGACATCTTT GGGGCCTACA GCATGGATGT GATTACTGGC ACATCATTTG
541 GAGTGAACAT CGACTCTCTC ACAAATCCAC AAGACCCCTT TGTGGAGAGC ACTAAGAAGT
601 TCCTAAAATT TGGTTTCTTA GATCCATTAT TTCTCTCAAT AATACTCTTT CCATTCTTA
661 CCCCAGTTTT TGAAGCATTAA AATGTCCTCTC TGTTTCCAAA AGATACCATA AATTTTTTAA
721 GTAAATCTGT AAACAGAAATG AAGAAAAGTC GCCTCAACGA CAAACAAAAAG CACCGACTAG
781 ATTTCTTCA GCTGATGATT GACTCCCAGA ATTGAAAGA AACTGAGTCC CACAAAGCTC
841 TGTCTGATCT GGAGCTCGCA GCCCAGTCAA TAATCTTCAT TTTGCTGGC TATGAAACCA
901 CCAGCAGTGT TCTTCCTTC ACTTTATATG AACTGGCCAC TCACCCGTAT GTCCAGCAGA
961 AACTGAAAAA GGAGATTGAT GCAAGTTTGC CCAATAAGGC ACCACCTTAC TATGATGCCG
1021 TGGTACAGAT GGAGTACCTT GACATGGTGG TGAATGAAAC ACTCAGATTA TTCCCAGTTG
1081 CTATTAGACT TGAGAGGACT TGCAAGAAAATGTTGAAAT CAATGGGGTA TTCAATTCCA
1141 AAGGGTCAAT GGTGGTGATT CCAACTTATG CTCTTCACCA TGACCCAAAG TACTGGACAG
1201 AGCCTGAGGA GTTCCGCCCCCT GAAAGGTTCA GTAAGAAGAA GGACAGCATA GATCCTTACA
1261 TATACACACC CTTTGGAACT GGACCCAGAA ACTGCATTGG CATGAGGTTT GCTCTCATGA
1321 ACATGAAACT TGCTCTAATC AGAGTCCTTC AGAACCTCTC CTTCAAACCT TGTAAAGAAA
1381 CACAGATCCC CTTGAAATTAA GACACGCAAG GACTTCTCA ACCAGAAAAA CCCATTGTT
1441 TAAAGGTGGA TTCAAGAGAT GGAACCCCTAA GTGGAGAAc a tcaccatcac catcacTGAg
1501 atctgcag

```

## SEQUENCE LISTING

<110> EPIDAUROS Biotechnologie AG  
<120> Identification of the genetic determinants of the polymorphic CYP3A5 expression  
<130> E 3103 PCT  
<140>  
<141>  
  
<150> EP 00 12 8627.7  
<151> 2000-12-28  
  
<150> EP 01 10 0172.4  
<151> 2001-01-16  
  
<150> EP 01 11 8884.4  
<151> 2001-08-16  
  
<150> US 60/258,684  
<151> 2000-12-28  
  
<150> US 60/258,952  
<151> 2000-12-29  
  
<150> US 60/262,859  
<151> 2001-01-18  
  
<150> US 60/312,825  
<151> 2001-08-16  
  
<160> 239  
  
<170> PatentIn Ver. 2.1  
  
<210> 1  
<211> 21  
<212> DNA  
<213> Homo sapiens  
  
<400> 1  
acaggcacag aaacccacaa g  
  
<210> 2  
<211> 19  
<212> DNA  
<213> Homo sapiens  
  
<400> 2  
atcgccacctt gccttcttc  
  
<210> 3  
<211> 19  
<212> DNA  
<213> Homo sapiens  
  
<400> 3  
ccctgcttcg gcttgtca

21

19

19

<210> 4  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 4  
cacagcctgc tttatggc atga 24

<210> 5  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 5  
gatccttgtt aggacaagcc t 21

<210> 6  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 6  
caagcactga tttgggtcact tcct 24

<210> 7  
<211> 22  
<212> DNA  
<213> Homo sapiens

<400> 7  
gggatgggac cgtaagtggaa ac 22

<210> 8  
<211> 26  
<212> DNA  
<213> Homo sapiens

<400> 8  
taatcacatt ggagttctga caaatg 26

<210> 9  
<211> 30  
<212> DNA  
<213> Homo sapiens

<400> 9  
aaaaacacct tacaaaagta tcatcgata 30

<210> 10  
<211> 27  
<212> DNA  
<213> Homo sapiens

<400> 10  
cctacttaggt ctctgacttg gaaccat 27

<210> 11  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 11  
gccgagacgc accattacac t 21

<210> 12  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 12  
cacccatccc ttccccactca t 21

<210> 13  
<211> 28  
<212> DNA  
<213> Homo sapiens

<400> 13  
tgatggttcc aagttagaga ccttagtag 28

<210> 14  
<211> 35  
<212> DNA  
<213> Homo sapiens

<400> 14  
aattgttagac atctttctct taagtttaatt cccag 35

<210> 15  
<211> 25  
<212> DNA  
<213> Homo sapiens

<400> 15  
tctgcattgcc aacagtgaac aatct 25

<210> 16  
<211> 19  
<212> DNA  
<213> Homo sapiens

<400> 16  
ggcacgcacc agcatgtcc 19

<210> 17  
<211> 19  
<212> DNA  
<213> Homo sapiens

<400> 17  
ctggctgagt gccgtggct 19

<210> 18  
<211> 25  
<212> DNA  
<213> Homo sapiens

<400> 18  
tgagcgcttc atgtattctg gctat 25

<210> 19  
<211> 31  
<212> DNA  
<213> Homo sapiens

<400> 19  
aaatattttc aaagtcacac tctgacaaca g 31

<210> 20  
<211> 29  
<212> DNA  
<213> Homo sapiens

<400> 20  
taacaggatc tcatgctttt ttcatggct 29

<210> 21  
<211> 30  
<212> DNA  
<213> Homo sapiens

<400> 21  
caactccaata ttcacaatag ccactattca 30

<210> 22  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 22  
actcctacgt atcccttccaa gccc 24

<210> 23  
<211> 28  
<212> DNA  
<213> Homo sapiens

<400> 23  
gctaaggaa acaggcatag aaacttac 28

<210> 24  
<211> 18  
<212> DNA  
<213> Homo sapiens

<400> 24  
ggagcttccc tgccctgc 18

<210> 25  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 25  
tccttctcca gcacataaaat c 21

<210> 26  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 26  
aaattagaag gtggatggga g 21

<210> 27  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 27  
gagtaactca ccagccctct g 21

<210> 28  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 28  
aacacctcaga actccctccc a 21

<210> 29  
<211> 23  
<212> DNA  
<213> Homo sapiens

<400> 29  
gacatctctg aatacggtcc ttc 23

<210> 30  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 30  
gcacatagtt tataacggca a 21

<210> 31  
<211> 23  
<212> DNA  
<213> Homo sapiens

<400> 31  
agaacctaag gttgctgtgt gtc 23

<210> 32  
<211> 21  
<212> DNA  
<213> Homo sapiens  
  
<400> 32  
tgcaagatgt taccactggg c

21

<210> 33  
<211> 20  
<212> DNA  
<213> Homo sapiens  
  
<400> 33  
cgccccacat acactcagaa

20

<210> 34  
<211> 21  
<212> DNA  
<213> Homo sapiens  
  
<400> 34  
agaccattt taggaagctc g

21

<210> 35  
<211> 22  
<212> DNA  
<213> Homo sapiens  
  
<400> 35  
caagggtag tccactgagt tc

22

<210> 36  
<211> 18  
<212> DNA  
<213> Homo sapiens

<400> 36  
ctctttgag ttgcagcg

18

<210> 37  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 37  
aggtagtct aactcagctt g

21

<210> 38  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 38  
gacagctaaa gtgtgtgagg g

21

<210> 39  
<211> 21  
<212> DNA  
<213> Homo sapiens  
  
<400> 39  
aatgggttcc agttgagaat c

21

<210> 40  
<211> 21  
<212> DNA  
<213> Homo sapiens  
  
<400> 40  
attgttgtgc ctgatttcaa g

21

<210> 41  
<211> 19  
<212> DNA  
<213> Homo sapiens  
  
<400> 41  
agaagccata gggaggttg

19

<210> 42  
<211> 20  
<212> DNA  
<213> Homo sapiens  
  
<400> 42  
gactgtcctc caagcattct

20

<210> 43  
<211> 21  
<212> DNA  
<213> Homo sapiens  
  
<400> 43  
gatgccatga tgaggagtgt g

21

<210> 44  
<211> 19  
<212> DNA  
<213> Homo sapiens  
  
<400> 44  
accaggccca gcaatattg

19

<210> 45  
<211> 25  
<212> DNA  
<213> Homo sapiens  
  
<400> 45  
aaatacttca cgaatactat gatca

25

<210> 46  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 46  
cagggacata attgattatc tttg 24

<210> 47  
<211> 19  
<212> DNA  
<213> Homo sapiens

<400> 47  
tactggttgg gaggtggag 19

<210> 48  
<211> 23  
<212> DNA  
<213> Homo sapiens

<400> 48  
catgatgttc ttaatgctac agg 23

<210> 49  
<211> 25  
<212> DNA  
<213> Homo sapiens

<400> 49  
gaagaggtca agatacatgg tgtta 25

<210> 50  
<211> 23  
<212> DNA  
<213> Homo sapiens

<400> 50  
tgcacaacac tctacacaga ctc 23

<210> 51  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 51  
tgggcttgca a 11

<210> 52  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 52  
tgggcgtgca a 11

<210> 53  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 53  
gcatggtaa a 11

<210> 54  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 54  
gcatgagtaa a 11

<210> 55  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 55  
gggtgtgtg c 11

<210> 56  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 56  
gggtatgtg c 11

<210> 57  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 57  
tgtgcattc t 11

<210> 58  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 58  
tgtgcaattc t 11

<210> 59  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 59  
gcccccaccc t 11

<210> 60  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 60  
gccccgcctc c 11

<210> 61  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 61  
ctcacactgg g 11

<210> 62  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 62  
ctcacgctgg g 11

<210> 63  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 63  
gagacgcacc a 11

<210> 64  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 64  
gagacacacc a 11

<210> 65  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 65  
tgtgtgtggg a 11

<210> 66  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 66  
tgtgtatggg a 11

<210> 67  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 67  
atccatgtat a 11

<210> 68  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 68  
atccacgtat a 11

<210> 69  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 69  
catcttaccc c 11

<210> 70  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 70  
catctcaccc c 11

<210> 71  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 71  
tctattgcta t 11

<210> 72  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 72  
tctatacgta t 11

<210> 73  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 73  
ggcagggaaag c 11

<210> 74  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 74  
ggcagagaag c 11

<210> 75  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 75  
ccaggcaaac a 11

<210> 76  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 76  
ccaggttaaac a 11

<210> 77  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 77  
tcaaggagaa g 11

<210> 78  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 78  
tcaagaagta g 11

<210> 79  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 79  
gtacacatgg a 11

<210> 80  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 80  
gtacatatgg a 11

<210> 81  
<211> 10  
<212> DNA  
<213> Homo sapiens

<400> 81  
catggacttt 10

<210> 82  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 82  
catgggactt t 11

<210> 83  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 83  
gatagcaggc c 11

<210> 84  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 84  
gatagtaggc c 11

<210> 85  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 85  
agaatcgggc t 11

<210> 86  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 86  
agaatagggc t 11

<210> 87  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 87  
ttattctgtc t 11

<210> 88  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 88  
tttattatgtc t 11

<210> 89  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 89  
gtacaggaag g 11

<210> 90  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 90  
cctaaaaaat g 11

<210> 91  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 91  
tcttttatct t 11

<210> 92  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 92  
tctttgatct t 11

<210> 93  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 93  
gagtctgcac a 11

<210> 94  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 94  
gagtccgcac a 11

<210> 95  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 95  
gaagtcaaga a 11

<210> 96  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 96  
agtcgtcaag a 11

<210> 97  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 97  
aggaagtatt c 11

<210> 98  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 98  
aggaattatt c 11

<210> 99  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 99  
agagagcttc a 11

<210> 100  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 100  
agagaacttc a 11

<210> 101  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 101  
cttcaatagt a 11

<210> 102  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 102  
cttcagtagt a 11

<210> 103  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 103  
tccaaacttat g 11

<210> 104  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 104  
tccaaattat g 11

<210> 105  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 105  
cgaaaactaca t 11

<210> 106  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 106  
cgaaaattaca t 11

<210> 107  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 107  
aaggatttct a 11

<210> 108  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 108  
aaggacttct a 11

<210> 109  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 109  
agctccgttg t 11

<210> 110  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 110  
agctctgtt g t 11

<210> 111  
<211> 10  
<212> DNA  
<213> Homo sapiens

<400> 111  
cacctaccta 10

<210> 112  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 112  
caccttacct a 11

<210> 113  
<211> 830  
<212> DNA  
<213> Homo sapiens

<400> 113  
tggtcaccca ccatgtgtac agtaccctgc tagggtccag ggtcatgaaa gtaaataata 60  
ccagactgtg ccctttagga actcacctct gctaaggaa acaggcacag aaaccacaa 120  
gggtggtaga gagaaatag gacaatagga ctgtgtgagg gggataggag gcacccagag 180  
gaggaaatgg ttacatctgt tgaggaggt tggtaaggaa agacttaat agaaggggtc 240  
tgtctggctg ggcgtcaag gatgtgttagg agtcatctag gggcacaag tacactccag 300  
gcagagggaa ttgcatgggt aaagatctgc agttgtggct tgtggggatg gattcaagt 360  
attctggaat gaagacagcc atggaaacaa gggcaggtga gaggatattt aagaggcttc 420  
atgccaatgg ctccacttca gtttctgata agaactcagg ttccgtggac tccctgataa 480  
aactgattaa gttttagt attccccata gaatatgaac tcaaaggagg taagcaaagg 540  
ggtgtgtgcg attctttgc actggctgca gctcgagccc cgcctccttc tccagcacat 600  
aaacatttca gcagcttgac ctaagactgc tgtgcagggc agggatgctc cagggcagaca 660  
gcccagcaaa caacagcaca cagctgaaaag taagacttag aggagacagt tgaagaaggc 720  
aagtggcgat ggacctcatc ccaaatttgg cggtggaaac ctggcttctc ctggctgtca 780  
gcctgggtgc ctcttatctg tcagtaactg tccagattcc ttcctctgt 830

<210> 114  
<211> 830  
<212> DNA  
<213> Homo sapiens

&lt;400&gt; 114

tggtcaccca ccatgtgtac agtaccctgc tagggtccag ggtcatgaaa gtaaataata 60  
ccagactgtg cccttgagga actcacctct gctaaggaa acaggcacag aaaccacaa 120  
gggtggtaga gaggaaatag gacaatagga ctgtgtgagg gggataggag gcacccagag 180  
gaggaaatgg ttacatctgt gtgaggaggt tggtaaggaa agacttaat agaagggtc 240  
tgtctggctg ggcttgcaag gatgtgttagg agtcatctag gggcacaag tacactccag 300  
gcagaggaa ttgcatgagt aaagatctgc agtttggct tggatggatg gatttcaagt 360  
attctggaat gaagacagcc atggaaacaa gggcaggta gaggatattt aagaggcttc 420  
atgccaatgg ctccacttca gtttctgata agaactcagg ttccgtggac tccctgataa 480  
aactgattaa gtgtttatg attccccata gaatatgaac tcaaaggagg taagcaaagg 540  
ggtgtgtgcg attctttgt actggctgca gctcagcccc caccccttc tccagcacat 600  
aaacatttca gcagcttgac ctaagactgc tggcaggcc agggatgctc caggcagaca 660  
gcccgacaaa caacagcaca cagctgaaag taagactcg aggagacagt tgaagaaggc 720  
aagtggcgat ggacctcattt ccaaatttgg cggtggaaac ctggcttctc ctggctgtca 780  
gcctgggtgct cctctatctg tcagtaactg tccagattcc ttcctctgt 830

&lt;210&gt; 115

&lt;211&gt; 830

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 115

tggtcaccca ccatgtgtac agtaccctgc tagggtccag ggtcatgaaa gtaaataata 60  
ccagactgtg cccttgagga actcacctct gctaaggaa acaggcacag aaaccacaa 120  
gggtggtaga gaggaaatag gacaatagga ctgtgtgagg gggataggag gcacccagag 180  
gaggaaatgg ttacatctgt gtgaggaggt tggtaaggaa agacttaat agaagggtc 240  
tgtctggctg ggcttgcaag gatgtgttagg agtcatctag gggcacaag tacactccag 300  
gcagaggaa ttgcatgggt aaagatctgc agtttggct tggatggatg gatttcaagt 360  
attctggaat gaagacagcc atggaaacaa gggcaggta gaggatattt aagaggcttc 420  
atgccaatgg ctccacttca gtttctgata agaactcagg ttccgtggac tccctgataa 480  
aactgattaa gtgtttatg attccccata gaatatgaac tcaaaggagg taagcaaagg 540  
ggtgtgtgcg attctttgt actggctgca gctcagcccc caccccttc tccagcacat 600  
aaacatttca gcagcttgac ctaagactgc tggcaggcc agggatgctc caggcagaca 660  
gcccgacaaa caacagcaca cagctgaaag taagactcg aggagacagt tgaagaaggc 720  
aagtggcgat ggacctcattt ccaaatttgg cggtggaaac ctggcttctc ctggctgtca 780  
gcctgggtgct cctctatctg tcagtaactg tccagattcc ttcctctgt 830

&lt;210&gt; 116

&lt;211&gt; 830

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 116

tggtcaccca ccatgtgtac agtaccctgc tagggtccag ggtcatgaaa gtaaataata 60  
ccagactgtg cccttgagga actcacctct gctaaggaa acaggcacag aaaccacaa 120  
gggtggtaga gaggaaatag gacaatagga ctgtgtgagg gggataggag gcacccagag 180  
gaggaaatgg ttacatctgt gtgaggaggt tggtaaggaa agacttaat agaagggtc 240  
tgtctggctg ggcttgcaag gatgtgttagg agtcatctag gggcacaag tacactccag 300  
gcagaggaa ttgcatgggt aaagatctgc agtttggct tggatggatg gatttcaagt 360  
attctggaat gaagacagcc atggaaacaa gggcaggta gaggatattt aagaggcttc 420  
atgccaatgg ctccacttca gtttctgata agaactcagg ttccgtggac tccctgataa 480  
aactgattaa gtgtttatg attccccata gaatatgaac tcaaaggagg taagcaaagg 540  
ggtgtgtgcg attctttgt actggctgca gctcagcccc caccccttc tccagcacat 600  
aaacatttca gcagcttgac ctaagactgc tggcaggcc agggatgctc caggcagaca 660  
gcccgacaaa caacagcaca cagctgaaag taagactcg aggagacagt tgaagaaggc 720  
aagtggcgat ggacctcattt ccaaatttgg cggtggaaac ctggcttctc ctggctgtca 780  
gcctgggtgct cctctatctg tcagtaactg tccagattcc ttcctctgt 830

&lt;210&gt; 117

<211> 775  
<212> DNA  
<213> Homo sapiens

<400> 117

ggagtgacct gatttccag gtgctgtctg tcacccctt ctttgactag gaaaggaaac 60  
tccttgaccc ctgcgccttc tcaagtgagg caatgcctcg ccctgcttcg gcttgcac 120  
agcacgcgtc acccaactgtc ctgcacccac tgtctggcac tccctagtga gatgaacccg 180  
gtacctcaga tgaaaatgca gaaatcaccc gtcttctgtg taactcacgc tgggagctgt 240  
agaccggagc tgttcctatt cgcccatctt ggctccaccc cccgagtttt ggctttaat 300  
tgaaaagtgt a ttgatgtggg aaggagataa tgccatgtat ttatgagcac atattagagg 360  
gtctgagaca atgcatgtga taaaaggta cctaaggaag aaaaaaaagaaa caagggaaaga 420  
caactggaaag aacgtgatgc tgggagtc tgggcccacca aagtctggag aaaagtggta 480  
accacaaggc tcccagccta gtttcaactga agacccgtac actagggtgac ttatggatc 540  
cttggtagga caagccttga aaagttttac aatagcaa at gtggacgtt tcagaaccaa 600  
atgatgtcac gtgtgttattt gtgtgtgtgt gtcaactgtgt gtgtttaaaa atcatgacaa 660  
ataaagcagg ctgtgaagag gggattccca tgctcgtgtg cctgataaca caactatcac 720  
aaacgctttt cgaaaccaca agtttgcaca aaggaatcc caaccttaca caaaa 775

<210> 118  
<211> 1110  
<212> DNA  
<213> Homo sapiens

<400> 118

taatgaaaat aagaattatt ttgatggct taacagtgac atttatatca tctgttttat 60  
ctggagcatt ctataataag ttatattaa gcaaatcaat aaaaacctct tacaaaagta 120  
tcatcggtata cttcctgaa cattaaggag aaatctatag aactgaatga atgagaacca 180  
acaagtaaat atatgtgatc attgtAACCA ttgttgggtgt ggggcatttt tcagaactcc 240  
aatgtgatta ttaacatagg tgagaattaa tccactgtga ctttgccttgc tgcttagaaa 300  
gaacattcat agttaatta tgccctttt gaccaagcac agtggctcat gcctgtatc 360  
ccagcactt gggaggccga ggtgggtgga tcacctgagg tcaggagttc gagaccagcc 420  
tgaccaacat ggtgaaaccc catctctact aaaaatacaa aatattagcta ggtgtgg 480  
tatgcaccta taatctcagc tacccaggag gctgaggcag gagaatcaact tgaacctgga 540  
ggcagagggt gcagtgagcc gagacacacc attacactcc agcctgggtt acagagttag 600  
attccatctc aaaaaaaaaa aaaaaaaaaatt atgcctttt gaagcacata cattttataa 660  
catacaactg aatcccttat tatattatta gtttgattt aatgtttca aaccatctcc 720  
cctgatattt ctgggagatg gaaaacatgt tttcttacac ctcttgcatt ccattctcaa 780  
ctccccactg tcttactgca atgaacactt aataagaaac agtcaattgg tcaattgatt 840  
gggcaacagg ctaaacacac tcattccttgc tctttccca cttttttctt tactttccct 900  
tcctgagtaa cttatcctaa agtcattagg tgggtggcag ccagatgggt gccacacatt 960  
aaggtagaaa agagagtgtc atgatgggtc caagtcagag accttagtagg gtgaggatca 1020  
agtaggtgtt cacgtggaga aacagccccg cctgtgtgtg ggagtccaag caagcagaga 1080  
aaatgtcgac acagaggggt ggcctgaaaa 1110

<210> 119  
<211> 837  
<212> DNA  
<213> Homo sapiens

<400> 119

caaaaattag ctaggtgtgg tggtatgcac ctataatctc agtacccag gaggctgagg 60  
caggagaatc acttgaacct ggaggcagag gttgcagtga gcccggacgc accattacac 120  
tccagcctgg gtgacagagt gagattccat ctcaaaaaaaaa aaaaaaaaaaa attatgcctt 180  
tttgaagcac atacattttta taacatcaa ctgaatccct tattatatta ttatgtttga 240  
tttaatgttt tcaaaaccatc tcccctgata tttctggag atgggaaaca ttttttttta 300  
cacctcttgc attccattct caactcccaa ctgtcttact gcaatgaaca cttaaataaga 360  
aacagtcaat tggtaatttgcatttggcaac aggctaaaca cactcattcc ttgtctgttc 420  
ccacttctt ctttacttcc ctttccctgag taacttaccc taaagtcaat aggtgggtgg 480  
cagccagatg gtggccacac attaaggtag aaaaagaggt gtcatgtgg ttccaagtca 540  
gagaccttagt aggtgagga tcaagtaggt gttcacgtgg agaaacagcc cggcctgtgt 600

atgggagtcc aagcaagcag agaaaatgtc gacacagagg ggtggcctga ~~aaaaaycaycc~~ 666  
agagcctaaa cagggcatgg agaacatatt tagggcatga ggtgaggagg gcatccatga 720  
gtggaaaggg atgggtgagg tttcaactaca taaagggat tgatgaaata agtaaataaa 780  
gtatactgga agccaggtgt gtcactttt cagaaaagag tcatacgattc agaaagg 837

<210> 120  
<211> 1197  
<212> DNA  
<213> Homo sapiens

<400> 120  
ttctttctt actttccctt cctgagtaac ttatcctaaa gtcatttaggt gggtggcagc 60  
cagatgggg ccacacatta aggttagaaaa gagagtgtca tgatggttcc aagtcaagaga 120  
ccttagtaggg tgaggatcaa gtaggtgttc acgtggagaa acagccccc ctgtgtgtgg 180  
gagtccaaagc aagcagagaa aatgtcgaca cagaggggtg gcctgaaaaaa gcagccagag 240  
cctaaacagg gcatggagaa catattnagg gcatgagggtg aggagggcat ccatgagtgg 300  
gaagggatgg gtgagggttc actacataaa ggggattgtat gaaataagta aataaaagtat 360  
actggaagcc aggtgtgtca ctttgcaga aaagagtcat ggattcgaa agggagaaaa 420  
ctagcagaa tcctatgaaa tttagattaa atggatgtat ccacgttatat tcataccctt 480  
ctagatagat aaatggtag ataggtgata aaaagataac aagaggacaa gataattaga 540  
tagacataaa tgtatgtatg tggttggtg tggatcaaaa aaaacatata ctccctactt 600  
ctctccactg atagggctag gtaacaatgg catttcaata gcaatgagca cacttagtgg 660  
ccagatcttgc gcttattaaat accatttcc actgaaagga accagagctt ttttagagaaaa 720  
tggctgatcc cagggccagg attaagaatg ttcaagataa gccttaggata cattttgtgc 780  
caggaagcaa gaagatgttc aaatgatttc caagtaatgt ttggaaatga tatttggaaaa 840  
tgatttccaa atgatatttc caaatgattt ccaaatgata tatggaaaca cttaaagact 900  
ccactaaaga actatttagat ctgataaaaca aattcagtaa tggtgctgaa tacaaaatca 960  
acatacaaaa accagtagca tttctgcatg ccaacagtga acaatctggc aaaaataaaa 1020  
aatgtaatcc catttacaat aaccccaaat aaaactaaat acctggaaat taacttaaga 1080  
gaaagatgtc tacaattaaat attgtaaaaac actgtatgaaag gaaattgaaag aagacacaaa 1140  
aaagaaggat attccatgtt tatatattgt aagcattaaat attgttaaaa atgtcca 1197

<210> 121  
<211> 817  
<212> DNA  
<213> Homo sapiens

<400> 121  
caaagcaaaa atggacaaaat ggatcagatc aagtaaaaaaaa gcttctgtac cacaagaaaa 60  
gcaatcaaca aagtgaagac acaaaccaca gaatgggaga aaatattttc aaagtccacac 120  
tctgacaaca gattaatagc cagaatacat gaagcgctca aacaactctg taaggaaaaaa 180  
tctaataatc caatcaaaaa atggccaaaa ttgtataga cattttcaaa aagaagacat 240  
acaaatgccat cataggcata tgataaggtg ctcaacatca ctggcattt gagaatgca 300  
aatcaaaaacc acaatgagat atcatctcac cccagctaaa atggttttt tccaaaagac 360  
aggcaacaac aaatgccagc gagaatgtgg agaaaaggaa acccttgcactgttgg 420  
taaatttagtgc caaccactat agagaacaat ttggagggttc ctcaaaacat taaaattaac 480  
attnaaataga gctaccacaa tatccagaaaa tccccatgtc gggtatatac ctggaaagaaa 540  
ggaaatcata tattgaagag ataacatcac tccaaatattc acaatagccat ctattcaca 600  
atgccaagat ttggaaagca cctaagtgtc catcaacaga tgaatggata aagaaagtac 660  
tccaaattata cacaatggag cacaattcag ccatgaaaaaa agcatgagat cctgttatct 720  
gtaataatat ggatggaaact ggaggtcatc atgttaagtg aaataagccat ggcacagaaaa 780  
cacagatattt gcaagttctc acatacttgtt gggatct 817

<210> 122  
<211> 1126  
<212> DNA  
<213> Homo sapiens

<400> 122  
cattaaaatt aacattaaat agagctacca caatatccag aaatccccat gctgggtata 60

taccttggaaag aaaggaaatc atatattgaa gagataacat cactccaata 120  
 ccactattca caaatgc当地 gatttggaaag caacctaagt gtccatcaac agatgaatgg 180  
 ataaagaaag tactccaatt atacacaatg ggcacaatt cagccatgaa aaaagcatga 240  
 gatcctgtta tctgtataaa tatggatgga actggaggc atcatgttaa gtgaaataag 300  
 ccaggcacag aaacacagat attgcaaggc ctcacatact tgtggatct acaaataaaa 360  
 acaactgagc taatgtctgg gccttagtca gtgttgc当地 caagtactgg gagcacagct 420  
 tttaaaatac atcatgaatg cttaataca ggaatgaata gatgagaggc acaaactgg 480  
 tgggtgttct tctgatacac agtatcttcc ttgacagatt cagtacaact ctcacacagg 540  
 aagtctttc atgttatgtt accttatgag gaattaagtgc gagaacatg atttcttata 600  
 ttcccttg cagaacaaga ccaactttat tagttggac acagtgtggc tgcatttgag 660  
 tcccaagcaa ccattagtct atagctatca ccacagagtc agaggggatg agacgcccag 720  
 caatctcacc caagacaact ccaccaacat tcctgggtac ccacatgtg tacagtaccc 780  
 tgcttaggaac cagggtcatg aaagtaataa atacocagact gtgc当地tga ggagctcacc 840  
 tctgctaagg gaaacaggca tagaaactta caatgggtt agagagaaaa gaggacaata 900  
 ggactgtgtg agggggatag gaggcaccca gaggaggaaa tggttacatt tgggtgagga 960  
 ggttggtaag gaaaaattt agcagaaggc gtctgtctgg ctggcttgg aaggatacg 1020  
 aggagtcatc tagaggcac aggtacactc caggcagagg gaatttctgtg gttaaagatg 1080  
 tggtaggtgtg gcttggagg atggatttca attattcttag aatgaa 1126

<210> 123  
 <211> 624  
 <212> DNA  
 <213> Homo sapiens

<400> 123  
 gattaagctt ttcatgattc ctcatagaac atgaactcaa aagaggtcag caaagggtg 60  
 tggcgattc ttgttatttgc gctgc当地ta tagccctgcc tccttcttca gcacataaaat 120  
 ctttcagcag ctggctgaa gactgctgtg cagggcagag aagctccagg caaacagccc 180  
 agcaaacagc agcaactcagc taaaaggaag actcacagaa cacagttgaa gaaggaaaagt 240  
 ggcgatggac ctcatcccaa atttggcggt ggaaacctgg cttctcctgg ctgtcagcct 300  
 ggtgctcctc tatctgttag taactgttca aactcctctc ttgttccct tggacttggg 360  
 gtgctaatcg ggccccctt cccttatctg ttgttggat caaaagagat gttcaaggag 420  
 aagttagctga agtgttggac gctacaaaacg catagaagtt attattatct tatgcagatc 480  
 tatgaatgaa taaaataagca ttctcccat ccacccctta attttgggtga cttaggagggt 540  
 ttagggacacag catttggtag tggaaatgtat ttgatttagct tagatctgac gaagactaat 600  
 caatgaaaac atggcagcgg caga 624

<210> 124  
 <211> 624  
 <212> DNA  
 <213> Homo sapiens

<400> 124  
 gattaagctt ttcatgattc ctcatagaac atgaactcaa aagaggtcag caaagggtg 60  
 tggcgattc ttgttatttgc gctgc当地ta tagccctgcc tccttcttca gcacataaaat 120  
 ctttcagcag ctggctgaa gactgctgtg cagggcaggg aagctccagg taaacagccc 180  
 agcaaacagc agcaactcagc taaaaggaag actcacagaa cacagttgaa gaaggaaaagt 240  
 ggcgatggac ctcatcccaa atttggcggt ggaaacctgg cttctcctgg ctgtcagcct 300  
 ggtgctcctc tatctgttag taactgttca aactcctctc ttgttccct tggacttggg 360  
 gtgctaatcg ggccccctt cccttatctg ttgttggat caaaagagat gttcaaggag 420  
 aagttagctga agtgttggac gctacaaaacg catagaagtt attattatct tatgcagatc 480  
 tatgaatgaa taaaataagca ttctcccat ccacccctta attttgggtga cttaggagggt 540  
 ttagggacacag catttggtag tggaaatgtat ttgatttagct tagatctgac gaagactaat 600  
 caatgaaaac atggcagcgg caga 624

<210> 125  
 <211> 621  
 <212> DNA  
 <213> Homo sapiens

<400> 125  
gattaagctt ttcatgattc ctcatagaac atgaactcaa aagaggtcag caaagggtg 60  
tgtgcgattc tttgcatttg gctgcagta tagccctgcc tccttctcca gcacataaat 120  
ctttcagcag ctggctgaa gactgcttg cagggcaggg aagctccagg caaacagccc 180  
agcaaacagc agcaactcagc taaaaggaag actcacagaa cacagttgaa gaagggaaagt 240  
ggcgatggac ctcatccaa atttggcggt ggaaacctgg cttctcctgg ctgtcagcct 300  
ggtgctcctc tacatgttag taactgtcca aactcctctc ttgtttccct tggacttggg 360  
gtgctaatcg ggcccccttt cccttatctg tttgaagat caaaagagat gttcaagaag 420  
tagctgaagt gttggacgct acaaacgcat agaagttatt attatcttat gcagatctat 480  
gaatgaataa ataagcattt ctcccatcca ccttctaatt ttggtgacta ggaggttta 540  
gggacagcat ttgttagtgg gaatgattt attagcttag atctgacgaa gactaatcaa 600  
tggaaaacatg gcagcggcag a 621

<210> 126  
<211> 465  
<212> DNA  
<213> Homo sapiens

<400> 126  
agaaaaggac ctgatgagtg aatgcaatta ctgatgttgg agttgctgtt attattttatc 60  
gtgtacat tacctccctc tcttgaccat tccagttcct gaggtaactca ccagccctct 120  
gatctataaa gtcacaatcc ctgtgacctg atttctgttt cactttgttag atatgggacc 180  
cgtacatatg ggactttta agagacttgg aattccaggg cccacaccc tgcctttgtt 240  
gggaaatgtt ttgtcttac gtcaggtgag ttgcttgagc ttcttctttt gtttcttatg 300  
gttgc当地 aca tca gttttagt tccatcagta aaaatgcccc ttcttgggag ggagtctga 360  
gttccat tttcagaaat ggtggactg ggtgcagtg aggatctg taatctcagc 420  
ctctgtgagg ccaagactgg caaaattgtt gagcccaagga gttt 465

<210> 127  
<211> 33  
<212> PRT  
<213> Homo sapiens

<400> 127  
Met Asp Leu Ile Pro Asn Leu Ala Val Glu Thr Trp Leu Leu Leu Ala  
1 5 10 15

Val Ser Leu Val Leu Leu Tyr Leu Tyr Gly Thr Arg Thr Tyr Gly Thr  
20 25 30

Phe

<210> 128  
<211> 602  
<212> DNA  
<213> Homo sapiens

<400> 128  
catagacaag ggtgagtcct tcagtactta gagaaaattc aagagtgact taaaattccc 60  
cacttcaa atattctctg ttttcttgc tttcccttaa gacatctctg aatagcttcc 120  
ttcaactgcc agtggaaatg agtaggcctg atttcattgg acgcaactgt tttcagcccc 180  
aatttagaggt agggtttattt ctataaaaa taataatcaa ctgttattttt gtttctctc 240  
ccagggtctc tgaaaattt acacagagt ctataaaaag tatggaaaaa tgtgggggtg 300  
agtattctga aaacctccat tggatagacc tgctactgtg aggaggttac cccactgcag 360  
gatagtctct gcccaggctc tcatggatg aagcttctgt caacctaatac acaaacagag 420  
agaggtctc tgaaaagaaga ggataattac ttggagtag aatattgcaa tgggaatctg 480  
cttgcctta taaaactatgt gcaaatttcg ggagttaaac aagacaaaaga tgctccatag 540  
aaaatatgag aagaatctca taactgtttt gagataatta ttgttagcta caaagatcaa 600  
ta 602

<210> 129  
<211> 594  
<212> DNA  
<213> Homo sapiens

<400> 129  
cagtatcttccctgtttg gaccacatta cccttcata tatgaaggct tgggtggctc 60  
ctgtgtgaga ctcttgctgt gtgtcacacc ctaatgaact agaacctaag gttgctgtgt 120  
gtcgtaaac taggggtatg gattacataa cataatgatc aaagtctggc ttccctgggtg 180  
tggctccagc tgcaaatag ggctagtcaa gtttaatcag ctccgttgc cccacacaga 240  
acgttatgaag gtcaactccc tttgtgtggcc atcacagatc ccgacgtgat cagaacagt 300  
ctagtgaag aatgttatttc tttgtttcaca aatcgaaggg taagcatcca ttttttggaaa 360  
tttaaataat gattgatcca ctgattaaat ttttattttg aaaaaaaacat atattcacag 420  
aaggtaacct aaaaaatgtt caggaagggtt ccatgtactc ttcatcctgt cccgccccagt 480  
ggtaacatct tgcaatcttgc tatattgcaaa tatatatcta gtatattcat attatcaggt 540  
tggcacaaaaa gttaaaatgg caaactacag gctgggcata atggctcatg cctg 594

<210> 130  
<211> 594  
<212> DNA  
<213> Homo sapiens

<400> 130  
cagtatcttccctgtttg gaccacatta cccttcata tatgaaggct tgggtggctc 60  
ctgtgtgaga ctcttgctgt gtgtcacacc ctaatgaact agaacctaag gttgctgtgt 120  
gtcgtaaac taggggtatg gattacataa cataatgatc aaagtctggc ttccctgggtg 180  
tggctccagc tgcaaatcg ggctagtcaa gtttaatcag ctccgttgc cccacacaga 240  
acgttatgaag gtcaactccc tttgtgtggcc atcacagatc ccgacgtgat cagaacagt 300  
ctagtgaag aatgttatttc tttgtttcaca aatcgaaggg taagcatcca ttttttggaaa 360  
tttaaataat gattgatcca ctgattaaat ttttattttg aaaaaaaacat atattcacag 420  
aaggtaacct aaaaaatgtt caggaagggtt ccatgtactc ttcatcctgt cccgccccagt 480  
ggtaacatct tgcaatcttgc tatattgcaaa tatatatcta gtatattcat attatcaggt 540  
tggcacaaaaa gttaaaatgg caaactacag gctgggcata atggctcatg cctg 594

<210> 131  
<211> 594  
<212> DNA  
<213> Homo sapiens

<400> 131  
cagtatcttccctgtttg gaccacatta cccttcata tatgaaggct tgggtggctc 60  
ctgtgtgaga ctcttgctgt gtgtcacacc ctaatgaact agaacctaag gttgctgtgt 120  
gtcgtaaac taggggtatg gattacataa cataatgatc aaagtctggc ttccctgggtg 180  
tggctccagc tgcaaatcg ggctagtcaa gtttaatcag ctccgttgc cccacacaga 240  
acgttatgaag gtcaactccc tttgtgtggcc atcacagatc ccgacgtgat cagaacagt 300  
ctagtgaag aatgttatttc tttgtttcaca aatcgaaggg taagcatcca ttttttggaaa 360  
tttaaataat gattgatcca ctgattaaat ttttattttg aaaaaaaacat atattcacag 420  
aaggtaacct aaaaaatgtt caggaagggtt ccatgtactc ttcatcctgt cccgccccagt 480  
ggtaacatct tgcaatcttgc tatattgcaaa tatatatcta gtatattcat attatcaggt 540  
tggcacaaaaa gttaaaatgg caaactacag gctgggcata atggctcatg cctg 594

<210> 132  
<211> 60  
<212> PRT  
<213> Homo sapiens

<400> 132  
Met Trp Gly Thr Tyr Glu Gly Gln Leu Pro Val Leu Ala Ile Thr Asp

1 5 10 ..

Pro Asp Val Ile Arg Thr Val Leu Val Lys Glu Cys Tyr Tyr Val Phe  
20 25 30

Thr Asn Arg Arg Ser Leu Gly Pro Val Gly Phe Met Lys Ser Ala Ile  
35 40 45

Ser Leu Ala Glu Asp Glu Glu Trp Lys Arg Ile Arg  
50 55 60

<210> 133

<211> 601

<212> DNA

<213> Homo sapiens

<400> 133

cagtatcttc tccctgtttg gaccacatta cccttcata tatgaaggcct tgggtggctc 60  
ctgtgtgaga ctcttgcgt gtgtcacacc ctaatgaact agaacctaag gttgctgtgt 120  
gtcgtaacaac taggggtatg gattacataa cataatgate aaagtctggc ttccctgggtg 180  
tggctccagc tgcaaaatcg ggcttagtcaa gtttaatcag ctccgttgc cccacacaga 240  
acgtatgaag gtcaactccc tggctggcc atcacagatc cgcacgtgtat cagaacagt 300  
ctagtaaaag aatgttattc tggcttcaca aatcgaaggta taagcatcca ttttttggaaa 360  
tttaataat gattgatcca ctgattaaat ttttattttg aaaaaaaacat atattcacag 420  
aaggtaacct aaaaaatgtt cctaaaaaat gaagttcca tggacttcc atccctgtccc 480  
gcccagtggtaatcatttgc aatcttgat attgcaatat atatcttagt tattcatatt 540  
atcaggttgg cacaatggcaa actacaggct gggcataatg gtcatgcct 600  
g 601

<210> 134

<211> 603

<212> DNA

<213> Homo sapiens

<400> 134

agcggaaaac tcaaggaggt atgaaaataa gatgagtctt aatttagaaat gtaaaagaatg 60  
aatctgggaa caggtagaaa gtaagatcac agtccgttcc caaggggttag tccactgagt 120  
tcgagcttcc taaaaatggt ctttgcattt tatgtacaga aaagacatca caaaattcat 180  
tacaaaatgt cacttactgc tccatgcgtt agaaaagccat atccttctgg gacttgagtc 240  
tgcacatcta actacaggtt ctgatctgtt ttgtgtttagt atgttccccca tcattgccc 300  
gtatggagat gtattggta gaaacttgag gcgccaaagca gagaaggca agcctgtcac 360  
cttgcctccat ggagtcgaca gtcgcactgt tgggttactc cagtgaccag acaaaagcag 420  
ggcagcgcgtt caactccaaa gagccacca agaggagtg gtccttgcata ggcggcaagt 480  
cagcaaggaa aaaggccctt ctctcctgtt cacaggagcc aggattact tatctgttaa 600  
ctt 603

<210> 135

<211> 603

<212> DNA

<213> Homo sapiens ,

<400> 135

agcggaaaac tcaaggaggt atgaaaataa gatgagtctt aatttagaaat gtaaaagaatg 60  
aatctgggaa caggtagaaa gtaagatcac agtccgttcc caaggggttag tccactgagt 120  
tcgagcttcc taaaaatggt ctttgcattt tatgtacaga aaagacatca caaaattcat 180  
tacaaaatgt cacttactgc tccatgcgtt agaaaagccat atccttctgg gacttgagtc 240  
cgcacatcta actacaggtt ctgatctgtt ttgtgtttagt atgttccccca tcattgccc 300  
gtatggagat gtattggta gaaacttgag gcgccaaagca gagaaggca agcctgtcac 360  
cttgcctccat ggagtcgaca gtcgcactgt tgggttactc cagtgaccag acaaaagcag 420

tgccctgccat ggagtcgaca gtcgcactgt tgggttactc cagtgaccag ~~aaaaaaa~~ 400  
ggcagcgctg caactccaaa gagccaccta agagggagtg gctcccatga ggcggcaagt 540  
cagcaaggga aaagggcctt ctctcctgtg cacaggagcc aggatttact tatctgttaa 600  
ctt 603

<210> 136  
<211> 626  
<212> DNA  
<213> Homo sapiens

<400> 136  
attggacatg atagcttagat ttgtttcagg aaaacatcct gctttccaag gatttagatg 60  
aatgttttg ttcaactggtg actcaggtaa cacgtctca agaagccata gggaggttg 120  
gggagggaaag tcgtcaagaa gggaggttgaa ggactgcact ttgttattac ttctgacttc 180  
acgagtcaact ttctgccaaa gaaatctctc cttttgcctc tagcaccgac tagatttcct 240  
tcagctgatg attgactccc agaatttogaag agaaacttagg tcccacaaag gtaaccaagg 300  
agtgcctctg agggctactg gccccggacac taagagggag ggccttgc 360  
caggaagtat tccaggaaga tgagaatttt tgccacatag cagaacaaca cacattttaga 420  
tgttataaaat ggttagctgga ggcactttcc agaagcccac aggtatagcc atgttccagg 480  
ctgaaaggc aaccctaagc aaaccttagaa tgcttggagg acagtcaatg gtttggat 540  
cacctacatg agatcaaatg ccagttctca gccttcctca gatccaccaa gtgagaacct 600  
ctacttggaa atttatatca aacata 626

<210> 137  
<211> 623  
<212> DNA  
<213> Homo sapiens

<400> 137  
attggacatg atagcttagat ttgtttcagg aaaacatcct gctttccaag gatttagatg 60  
aatgttttg ttcaactggtg actcaggtaa cacgtctca agaagccata gggaggttg 120  
gggagggaaag tcaagaaggaa aggtttagga ctgcactttt gatttacttc tgacttcacg 180  
agtcaacttc tgccaaagaa atctctctt ttgttcttag caccgactag atttcctca 240  
gctgatgatt gactcccaga attcgaaaga aactgagtc cacaaggatg accaaggagt 300  
gcttctgagg gctactggcg gggacactaa gaggaggggc cttgttctga aaatgtgcag 360  
gaattattcc aggaagatga gaattttgc cacatagcag aacaacacac atttagatgt 420  
tataaaatggt agctggaggc actttccaga agccacagg tatagccatg ttccaggctg 480  
aaaggcaac cctaaagcaaa cctagaatgc ttggaggaca gtcagtgggt tggatcac 540  
ctacatgaga tcaaatgcca gttctcagcc tccctccatg ccaccaaggatg agaacctcta 600  
cttggaaatt tatataaacaata 623

<210> 138  
<211> 826  
<212> DNA  
<213> Homo sapiens

<400> 138  
agataaaatgta cttttaggat cattcaaggc acacacccat aacactgagt atgtaagaca 60  
gaaatgtct ctctggaaat tacagcagtg ctgggtctgg gatgccatga tgaggagtgt 120  
gtggccaca atcatgtaga ctttggaaa acctggatta aatgatttt gcgtcatct 180  
ggccctgtat aagatacata tcagaatgaa aaccactccc agtgtgactt tgaatttgc 240  
ttccatttt tctcttggg attagagaac ttcaacttgc ttcatctaa gctgtgatgt 300  
tgtacgttgc cctgatttac ctttttttgc ttctctctcc ttctcacttgc gtctgatctg 360  
gagctcgcag cccagtcaat aatcttcatt ttgttggct atgaaaccac cagcagtgtt 420  
ctttccatca ctttatatgaa actggccact caccctgtat tccaggcagaa actgcaaaag 480  
gagatttgcatg ctttttgc caataagggtt agggatgac ccctggagat gaaggaaaga 540  
ggtgaagcct tagaaaaat gcctcctcact cactccccag gagaattttt ataaaaagca 600  
taatcactga ttcttcaact gacataatgt aggaagcctc tgaggagaaa aacaaaggga 660  
gaaacataga gaacgggtgc tactggcaga agcataatgt ctttgc 720  
cctggttcact ctgtttactg ttatcacaat aatgtaatg aaaaaaaaaaaaaaaa 780

aaaaaaaaaa aggagtgtgg cgagaagatg gccaaacagg aacagg

020

<210> 139  
<211> 795  
<212> DNA  
<213> Homo sapiens

<400> 139  
gtccctgggg tgaggatgg cttgaatatac tcctacattc ataactcctc cacacatctc 60  
atgttagtac ttagcacatc aatggacatg ccaggattata aaatacttca cgaatactat 120  
gatcatttac cagttatgt tattctctgg agcttcta atcttcatgt tactgcattg 180  
actcagttga gagttattc aaaatctcag attatccaat tctgtttctt tccttccagg 240  
caccacatc ctatgtatgc gtggatcaga tggatcacct tgacatggtg gtgaatgaaa 300  
cactcagatt attcccgatg gctatttagac ttgagaggac ttgcaagaaa gatgttgaaa 360  
tcaatggggt atttccca aaagggtcaa tgggtgtat tccaacttat gctcttcacc 420  
atgacccaaa gtactggaca gggctgagg agttccccc tggaaaggta aagtctccag 480  
ggaaatggag ctcaaccctga cccaggctgg ttcaagcata ttctgcctot cttaatctac 540  
atgacaatcg tgggttgta caatcatttgc ttgttaagtc tttttatcac aaaaaagtga 600  
taattatcaa actttacaaa ccacagacta gaaaaaacga aactacatcc atccacagtc 660  
ccagcacaag acaaagataa tcaattatgt ccctgtggc atttttctac gcctatatag 720  
attttaaaa attagaatgg tatcaactttt tattttgggtt gaattgctgc ttacttgatt 780  
taacaggaaa ctatc 795

<210> 140  
<211> 796  
<212> DNA  
<213> Homo sapiens

<400> 140  
gtccctgggg tgaggatgg cttgaatatac tcctacattc ataactcctc cacacatctc 60  
atgttagtac ttagcacatc aatggacatg ccaggattata aaatacttca cgaatactat 120  
gatcatttac cagttatgt tattctctgg agcttcta atcttcatgt tactgcattg 180  
actcagttga gagttattc aaaatctcag attatccaat tctgtttctt tccttccagg 240  
caccacatc cctatgtatgc cgtggatcag atggaggatcc ttgacatggg ggtgaatgaa 300  
acactcagat tattcccgatg tgctatttaga cttgagagga ttgcaagaaa agatgttgaa 360  
atcaatgggg tattcattcc caaagggtca atgggtgtat ttccaaactta tgctcttcac 420  
catgacccaa agtactggac agaggctgag gagtccgccc ctgaaaggta caagtctcca 480  
ggaaatggga gctaccctg acccaggctg gttcaagcat attctgcctc tcttaatcta 540  
catgacaatc gtgtgggtgt acaatcattt gctttaatgt ctttttatca caaaaaagtgg 600  
ataattatca aactttacaaa accacagact agaaaaaaacg aaactacatc catccacagt 660  
ccagcacaag gacaaagata atcaattatgt tccctgtggc catttttctca cgcctatata 720  
gattttaaaa aattagaatgg gtatcaactttt ttattttgggtt tgaattgctgc ttacttgat 780  
ttaacaggaa actatc 796

<210> 141  
<211> 59  
<212> PRT  
<213> Homo sapiens

<400> 141  
Ala Leu Ser Asp Leu Glu Leu Ala Ala Gln Ser Ile Ile Phe Ile Phe  
1 5 10 15

Ala Gly Tyr Glu Thr Thr Ser Ser Val Leu Ser Phe Thr Leu Tyr Glu  
20 25 30

Leu Ala Thr His Pro Asp Val Gln Gln Lys Leu Gln Lys Glu Ile Asp  
35 40 45

Ala Val Leu Pro Asn Lys Ala Pro Pro Tyr Leu

50

55

<210> 142  
<211> 795  
<212> DNA  
<213> Homo sapiens

<400> 142  
gtccctggg tgaggatgg tttgaatata tcctacattt ataactcctc cacacatctc 60  
agttagtcac tgagcacatc aatggacatg ccagttatta aaatacttca cgaatactat 120  
gatcatttac cagttatgt tattctctgg agcttcta atctcaatag tactgcattg 180  
actcaggta gaggtaattt aaaatcttag attatccat tctgttctt tccttcagg 240  
caccacccatc ctatgtatgcc gtggtacaga tggagttaccc tgacatgggt gtgaatgaaa 300  
caactcaggatt attccctagg tctttagac ttgagaggac ttgcaagaaaa gatgttgaaa 360  
tcaatgggtt attcattccc aaagggtcaa ttgggtgtat tccaaacttat gctcttcacc 420  
atgaccctaaa gtactggaca gaggctgagg agttccgccc tgaaaggta aagtctccag 480  
ggaaatggag ctacccctga cccaggctgg ttcaagcata ttctgcctct cttaatctac 540  
atgacaatcg tgggttgta caatcatgtt ttgttaagtc tttttatcac aaaaaagtga 600  
taattatcaa actttacaaa ccacagacta gaaaaaacga aattacatcc atccacagtc 660  
ccagcacaag acaaagataa tcaattatgt ccctgtggc atttttctac gcctatata 720  
attttaaaa attagaatgg tatcactttt tattttttt gaattgtgc ttacttgatt 780  
taacaggaaa ctatc 795

<210> 143  
<211> 616  
<212> DNA  
<213> Homo sapiens

<400> 143  
caggcctggc acagagtcag tgctccataa atattttgtt aaacgatggg tggtgagtgc 60  
ttttactatc cagttttac ccagcttata gattaagtat gaagaggta agatacatgg 120  
tgttaagagt cgttttata tgcttcaaa gcattttgtt catattttt ctactttgct 180  
tccatctttt cttctttcac ttcatattt aatttccat atgcttggg aactattgt 240  
gatccccctt aaatttagaca cgcaaggact tcttcaacca gaaaaaccca ttgttctaaa 300  
ggtggattca agagatggaa ccctaagtgg agaatgagtt attctaaggaa cttctacttt 360  
ggtcttcaag aaagctgtgc cccagaacac cagagatttc aacttagtca ataaaaaccc 420  
gaaataaaga tgggtttaat ctaatgtact gcatgagtag ttgggtgattt tgtacattca 480  
ttgagctctc ccagagtctg tggtaggtgt tggcattat gtagtataaa ggaggtgacc 540  
aggtaagtga cagataggta gactcagctt ctctgcttct cataggacta cctctaccca 600  
cctctagttt gcatta 616

<210> 144  
<211> 1508  
<212> DNA  
<213> Homo sapiens

<400> 144  
ccatggctct gttttagca gttttctgg tgctccctta tctatatggg acccgtaac 60  
atggactttt taagagactg ggaattccag gcccacacc tctgccttgg ttggaaatg 120  
ttttgtccta tcgtcagggt ctctggaaat ttgacacaga gtgctataaa aagtatggaa 180  
aaatgtgggg aacgtatgaa ggtcaactcc ctgtgctggc catcacagat cccgacgtga 240  
tcagaacagt gctagtggaa gaatgttatt ctgtcttacca aatcgaagg tcttttagcc 300  
cagtgggatt tatggaaaagt gccatcttt tagctgagga tgaagaatgg aagagaatac 360  
ggtcatgtct gtcctcaacc ttccacagcg gaaaactcaa ggagatgttc cccatattt 420  
cccagtatgg agatgttattt gtgagaaact tgaggcgggaa agcagagaaa ggcaagcctg 480  
tcacctgaa agacatctt ggggcctaca gcatggatgtt gattactggc acatcattt 540  
gagtgaacat cgactcttc aacaatccac aagacccctt tggggagagc actaagaagt 600  
tcctaaaattt tgggttctta gatccattt ttctctcaat aatactctt ccattcctta 660  
ccccaggattt tgaagcattt aatgtcttc tgggtccaaa agataccata aattttttaa 720  
gtaaatctgtt aaacagaatg aagaaaaatgc gcctcaacga caaacaaaag caccgactag 780

atttccttca gctgatgatt gactcccaga attcgaaaga aactgagtcc cacaaagctc 840  
tgtctgatct ggagctcgca gcccagtcaa taatcttcat ttttgctggc tatgaaaacca 900  
ccagcagtgt tctttccttc acttttatatg aactggccac tcaccctgat gtccaggaga 960  
aactgaaaaa ggagattgat gcagtttgc ccaataaggc accacctacc tatgatgccg 1020  
tggtagatggatggat gacatggtgg tgaatgaaac actcagatta ttcccagttg 1080  
ctattagact tgagaggact tgcaagaaag atggttaaat caatgggta ttcattccca 1140  
aagggtcaat ggtgggtgatt ccaacttatg ctcttcacca tgacccaaag tactggacag 1200  
agcctgagga gttccgcct gaaaggttca gtaagaagaa ggacagcata gatccttaca 1260  
tatacacacc ctttggaaact ggaccaggaa actgcattgg catgaggtt gctctcatga 1320  
acatgaaact tgctctaatac agagtccctc agaacttctc cttcaaacct tggtaaagaaa 1380  
cacagatccc cttgaaatta gacacgcaag gacttcttca accagaaaaa cccatttttc 1440  
taaaggttga ttcaagagat ggaaccctaa gtggagaaca tcaccatcac catcactgag 1500  
atctgcag 1508

<210> 145  
<211> 60  
<212> PRT  
<213> Homo sapiens

<400> 145  
Pro Val Ala Ile Arg Leu Glu Arg Thr Cys Lys Lys Asp Val Glu Ile  
1 5 10 15

Asn Gly Val Phe Ile Pro Lys Gly Ser Met Val Val Ile Pro Asn Tyr  
20 25 30

Ala Leu His His Asp Pro Lys Tyr Trp Thr Glu Pro Glu Glu Phe Arg  
35 40 45

Pro Glu Arg Phe Ser Lys Lys Asp Ser Ile Asp  
50 55 60

<210> 146  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 146  
tttcagatc t 11

<210> 147  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 147  
tttcaatatc t 11

<210> 148  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 148  
tttaatagaa g 11

<210> 149  
<211> 11

<212> DNA  
<213> Homo sapiens

<400> 149  
tttaacagaa g

11

<210> 150  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 150  
attccccata g

11

<210> 151  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 151  
attcctcata g

11

<210> 152  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 152  
tagaatatga a

11

<210> 153  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 153  
tagaacatga a

11

<210> 154  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 154  
gtaacttac c

11

<210> 155  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 155  
gtaacgtatc c

11

<210> 156  
<211> 11

<212> DNA  
<213> Homo sapiens

<400> 156  
ttcacgtgga g

11

<210> 157  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 157  
ttcacatggaa g

11

<210> 158  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 158  
tggacgcAAC t

11

<210> 159  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 159  
tggacacaAC t

11

<210> 160  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 160  
gaggataATT a

11

<210> 161  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 161  
gaggAAAATT a

11

<210> 162  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 162  
catcattGCC C

11

<210> 163  
<211> 11

<212> DNA  
<213> Homo sapiens

<400> 163  
catcaactgcc c

11

<210> 164  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 164  
cagtcgcact g

11

<210> 165  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 165  
cagtcacact g

11

<210> 166  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 166  
actaagaagt t

11

<210> 167  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 167  
actaaaaagt t

11

<210> 168  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 168  
gatccattat t

11

<210> 169  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 169  
gatccgttat t

11

<210> 170  
<211> 11

<212> DNA  
<213> Homo sapiens

<400> 170  
caattccatt g 11

<210> 171  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 171  
caatttcatt g 11

<210> 172  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 172  
tgtcaatcta g 11

<210> 173  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 173  
tgtcagtcta g 11

<210> 174  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 174  
ttgtttgttt t 11

<210> 175  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 175  
ttgttcgttt t 11

<210> 176  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 176  
aaataaaagaa g 11

<210> 177  
<211> 11

<212> DNA  
<213> Homo sapiens

<400> 177  
aaatacagaa g

11

<210> 178  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 178  
tttgcgtcat c

11

<210> 179  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 179  
tttgcatcat c

11

<210> 180  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 180  
ttgacacctat t

11

<210> 181  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 181  
ttgacttgat t

11

<210> 182  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 182  
tattgttagat c

11

<210> 183  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 183  
tattgcagat c

11

<210> 184  
<211> 11

<212> DNA  
<213> Homo sapiens

<400> 184  
cctttccct t

11

<210> 185  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 185  
ccttccccct t

11

<210> 186  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 186  
cttatgcaga t

11

<210> 187  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 187  
cttatacaga t

11

<210> 188  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 188  
aatacggtca t

11

<210> 189  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 189  
aatacagtca t

11

<210> 190  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 190  
ggaggtatga a

11

<210> 191  
<211> 11

<212> DNA  
<213> Homo sapiens

<400> 191  
ggaggcatga a 11

<210> 192  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 192  
agtccgttc c 11

<210> 193  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 193  
agtccatttc c 11

<210> 194  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 194  
tctgccaag a 11

<210> 195  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 195  
tctgcgaaag a 11

<210> 196  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 196  
acccattgtt c 11

<210> 197  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 197  
acccactgtt c 11

<210> 198  
<211> 11

<212> DNA  
<213> Homo sapiens

<400> 198  
gctactggct g 11

<210> 199  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 199  
gctaccggct g 11

<210> 200  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 200  
gaaatcaccc g 11

<210> 201  
<211> 11  
<212> DNA  
<213> Homo sapiens

<400> 201  
gaaattaccc g 11

<210> 202  
<211> 33  
<212> DNA  
<213> Homo sapiens

<400> 202  
gctctactgt catttctaac cataatctct tta 33

<210> 203  
<211> 27  
<212> DNA  
<213> Homo sapiens

<400> 203  
gcttcataatg atgaagggtatgtgg 27

<210> 204  
<211> 18  
<212> DNA  
<213> Homo sapiens

<400> 204  
tgtctttcag tatctctt 18

<210> 205  
<211> 19

<212> DNA  
<213> Homo sapiens

<400> 205  
tgtctttcaa tatctcttc 19

<210> 206  
<211> 444  
<212> DNA  
<213> Homo sapiens

<400> 206  
cattttagcc ttgtgagcac ttgatgattt acctgccttc aattttcac tgacctaata 60  
ttcttttga taatgaagta ttttaaacat ataaaacatt atggagagt gcataaggaga 120  
taccacgttgttaccaccc agcttaacga atgcctact gtcatttcta accataatct 180  
ctttaaagag ctctttgtc ttcaatatac tcttccctgt ttggaccaca ttacccttca 240  
tcatatgaag ccttgggtgg ctccctgtgt agactcttgc tgtgtgtcac accctaata 300  
actagaacct aaggttgtgt tgcgtac aactagggtt atggattaca taacataatg 360  
atcaaagtct ggcttccctgg gtgtggctcc agctgcagaa tcgggctagt gaagtttaat 420  
cagctccgtt gtccccacac agaa 444

<210> 207  
<211> 830  
<212> DNA  
<213> Homo sapiens

<400> 207  
tggtcaccca ccatgtgtac agtaccctgc tagggtccag ggtcatgaaa gtaaataata 60  
ccagactgtg cccttgagga actcacctct gctaaggaa acaggcacag aaaccacaa 120  
gggtggtaga gaggaaatag gacaatagga ctgtgtgagg gggataggag gcaccacag 180  
gaggaaatgg ttacatctgt gtgaggaggt tggtaaggaa agacttaac agaagggtc 240  
tgtctggctg ggcttgcag gatgtgttagg agtcatctag gggcacaag tacactccag 300  
gcagaggaa ttgcattgggt aaagatctgc agttgtggct tggggatg gatttcaagt 360  
attctggaat gaagacagcc atggaaacaa gggcagggtga gaggatattt aagaggcttc 420  
atgcaatgg ctccacttca gtttctgata agaactcagg ttccgtggac tccctgataa 480  
aactgattaa gtttttatg attccctata gaatatgaac tcaaaggagg taagcaaagg 540  
ggtgtgtcg attctttgtc actggctgc gctgcagccc cacctccttc tccagcacat 600  
aaacatttca gcagcttgcac ctaagactgc tgcagggtggc agggatgc tggcagaca 660  
gcccagcaaa caacagcaca cagctgaaag taagacttag agggagacagt tgaagaaggc 720  
aagtggcgat ggacctcattt ccaaatttgg cggtggaaac ctggcttctc ctggctgtca 780  
gcctgggtgtc cctctatctg tcaactg tccagattcc ttcctctgt 830

<210> 208  
<211> 830  
<212> DNA  
<213> Homo sapiens

<400> 208  
tggtcaccca ccatgtgtac agtaccctgc tagggtccag ggtcatgaaa gtaaataata 60  
ccagactgtg cccttgagga actcacctct gctaaggaa acaggcacag aaaccacaa 120  
gggtggtaga gaggaaatag gacaatagga ctgtgtgagg gggataggag gcaccacag 180  
gaggaaatgg ttacatctgt gtgaggaggt tggtaaggaa agacttaat agaagggtc 240  
tgtctggctg ggcttgcag gatgtgttagg agtcatctag gggcacaag tacactccag 300  
gcagaggaa ttgcattgggt aaagatctgc agttgtggct tggggatg gatttcaagt 360  
attctggaat gaagacagcc atggaaacaa gggcagggtga gaggatattt aagaggcttc 420  
atgcaatgg ctccacttca gtttctgata agaactcagg ttccgtggac tccctgataa 480  
aactgattaa gtttttatg attccctata gaatatgaac tcaaaggagg taagcaaagg 540  
ggtgtgtcg attctttgtc actggctgc gctgcagccc cacctccttc tccagcacat 600  
aaacatttca gcagcttgcac ctaagactgc tgcagggtggc agggatgc tggcagaca 660  
gcccagcaaa caacagcaca cagctgaaag taagacttag agggagacagt tgaagaaggc 720

aagtggcgat ggacctcatc ccaaatttgg cggtggaaac ctggcttctc ctggctgtca 780  
gcctggtgct cctctatctg tcagtaactg tccagattcc ttcctctgt 830

<210> 209  
<211> 830  
<212> DNA  
<213> Homo sapiens

<400> 209  
tggtcaccca ccatgtgtac agtaccctgc tagggccag ggtcatgaaa gtaaataata 60  
ccagactgtg cccttgagga actcacctct gctaaggaa acaggcacag aaaccacaa 120  
gggtggtaga gaggaaatag gacaatagga ctgtgtgagg gggataggag gcaccagag 180  
gaggaaatgg ttacatctgt gtgaggaggt tggttaaggaa agactttaat agaaggggtc 240  
tgtctggctg ggcttgcaag gatgtgttagg agtcatctag ggggcacaag tacactccag 300  
gcagaggaa ttgcatgggt aaagatctgc agttgtggct tggggatgt gatttcaagt 360  
attctgaaat gaagacagcc atggaaacaa gggcaggtga gaggatattt aagaggcttc 420  
atgccaatgg ctccacttca gtttctgata agaactcagg ttccgtggac tccctgataa 480  
aactgatcaa gttgtttatg attccccata gaacatgaac tcaaaggagg taagcaaagg 540  
ggtgtgtcg attctttgtc actggctgca gctgcagccc cacctccttc tccagcacat 600  
aaacatttca gcagacttgc ctaagactgc tggcaggc agggatgctc caggcagaca 660  
gcccagcaa caacagcaca cagctgaaag taagactcg aggagacagt tgaagaaggc 720  
aagtggcgat ggacctcatc ccaaatttgg cggtggaaac ctggcttctc ctggctgtca 780  
gcctggtgct cctctatctg tcagtaactg tccagattcc ttcctctgt 830

<210> 210  
<211> 837  
<212> DNA  
<213> Homo sapiens

<400> 210  
caaaaattag cttaggtgtgg tggtatgcac ctataatctc agctacccag gaggctgagg 60  
caggagaatc acttgaacct ggaggcagag gttcagtgta gcccagacgc accattacac 120  
tccagcctgg gtgacagagt gagattccat ctcaaaaaaaaa aaaaaaaaaa attatgcctt 180  
tttgaagcac atacattttta taacatacaa ctgaatccct tattatatta ttagtttga 240  
tttaatgttt tcaaaccatc tcccctgata tttctggag atggggaaaca tggggatcta 300  
cacctcttgc attccattct caactcccaa ctgtcttact gcaatgaaca cttaataaga 360  
aacagtcaat tggcaatttgg attgggcaac aggctaaaca cactcattcc ttgtctgttc 420  
ccacttcttt ctttactttc ctttcctgag taacgtatcc taaagtctt aggtgggtgg 480  
cagccagatg gtggcccacac attaaggtag aaaagagagt gtcatgtatgg ttccaagtca 540  
gagaccttagt agggtgagga tcaagtaggt gttcacgtgg agaaacagcc cggcctgtgt 600  
gtggggatcc aagcaaggcag agaaaaatgtc gacacagagg ggtggcctga aaaagcagcc 660  
agagcctaaa cagggcatgg agaacatatt tagggcatga ggtgaggagg gcatccatga 720  
gtggggaggg atgggtgagg ttctactaca taaagggat tggatgaaata agtaaataaaa 780  
gtataactgga agccaggtgt gtcacttttgc cagaaaaagag tcatggattc agaaagg 837

<210> 211  
<211> 837  
<212> DNA  
<213> Homo sapiens

<400> 211  
caaaaattag cttaggtgtgg tggtatgcac ctataatctc agctacccag gaggctgagg 60  
caggagaatc acttgaacct ggaggcagag gttcagtgta gcccagacgc accattacac 120  
tccagcctgg gtgacagagt gagattccat ctcaaaaaaaaa aaaaaaaaaa attatgcctt 180  
tttgaagcac atacattttta taacatacaa ctgaatccct tattatatta ttagtttga 240  
tttaatgttt tcaaaccatc tcccctgata tttctggag atggggaaaca tggggatcta 300  
cacctcttgc attccattct caactcccaa ctgtcttact gcaatgaaca cttaataaga 360  
aacagtcaat tggcaatttgg attgggcaac aggctaaaca cactcattcc ttgtctgttc 420  
ccacttcttt ctttactttc ctttcctgag taacttaccc taaagtctt aggtgggtgg 480  
cagccagatg gtggcccacac attaaggtag aaaagagagt gtcatgtatgg ttccaagtca 540

gagacctagt agggtgagga tcaagtaggt gttcacatgg agaaacagcc -,-,-,-,-,-  
gtgggagtc aagcaagcag agaaaaatgtc gacacagagg gttggcctga aaaagcagcc 660  
agagcctaaa cagggcatgg agaacatatt tagggcatga gttgaggagg gcatccatga 720  
gtgggaaggg atgggtgagg tttcaactaca taaagggat ttttgcataaataa 780  
gtatactgga agccaggtgt gtcactttt cagaaaagag tcacggattc agaaaagg 837

<210> 212  
<211> 602  
<212> DNA  
<213> Homo sapiens

<400> 212  
catagacaag ggtgagtcct tcagtagtta gagaaaattc aagagtgact taaaattccc 60  
cacttcaaat atattctctg ttttcttgc ttcccttaa gacatctctg aatagcttcc 120  
ttcaactgcc agtggaaagat agcaggccctg atttcattgg acacaactgtt tttcagcccc 180  
aatttagaggt agggtttattt ctatTTAAA taataatcaa cttgtatTTT gtttcccttc 240  
ccagggtctc tgaaattttg acacagagtg ctatTTAAAG tatggaaaaaa ttttgggggtg 300  
agtattctga aaacctccat tggatagacc tgctactgtg aggagggttac cccactgcag 360  
gatagtctct gcccagggtct tcattggatg aagcttctgt caacctaaat acaaacagag 420  
agaggttctc tgaaagaaga ggataattac ttgggagtag aatattgcaa tggaaatctg 480  
cttgcctta taaactatgt gcaaattcag ggaggtaaac aagacaaaga tgctccatag 540  
aaaatatgag aagaatctca taactgtttt gagataatta ttgttagcta caaagatcaa 600  
ta 602

<210> 213  
<211> 602  
<212> DNA  
<213> Homo sapiens

<400> 213  
catagacaag ggtgagtcct tcagtagtta gagaaaattc aagagtgact taaaattccc 60  
cacttcaaat atattctctg ttttcttgc ttcccttaa gacatctctg aatagcttcc 120  
ttcaactgcc agtggaaagat agcaggccctg atttcattgg acgcactgtt tttcagcccc 180  
aatttagaggt agggtttattt ctatTTAAA taataatcaa cttgtatTTT gtttcccttc 240  
ccagggtctc tgaaattttg acacagagtg ctatTTAAAG tatggaaaaaa ttttgggggtg 300  
agtattctga aaacctccat tggatagacc tgctactgtg aggagggttac cccactgcag 360  
gatagtctct gcccagggtct tcattggatg aagcttctgt caacctaaat acaaacagag 420  
agaggttctc tgaaagaaga ggaaaattac ttgggagtag aatattgcaa tggaaatctg 480  
cttgcctta taaactatgt gcaaattcag ggaggtaaac aagacaaaga tgctccatag 540  
aaaatatgag aagaatctca taactgtttt gagataatta ttgttagcta caaagatcaa 600  
ta 602

<210> 214  
<211> 603  
<212> DNA  
<213> Homo sapiens

<400> 214  
agcggaaaac tcaaggaggt atggaaataa gatgagtctt aatttagaaat gtaaagaatg 60  
aatctgggaa caggtagaaa gtaagatcac agtccgttca caagggttag tccactgagt 120  
tcgagctcc taaaaatgtt cttttatctt tatgtacaga aaagacatca caaaattcat 180  
tacaaaatgt cactactgc tccatgctgg agaaagccat atccttctgg gacttgagt 240  
tgcacattt actacaggta ctgatctgtt ttgtgcttag atgttccccca tcactgccc 300  
gtatggagat gtattggta gaaaacttgag gcgggaagca gagaaaggca agcctgtcac 360  
cttggaaagag taatgttaggag cacagccatg gggttcttag ctgtcatgag cccttccagc 420  
tgcctgcctt ggagtgcaca gtcgcactgt tgggttactc cagtgaccag acaaaagcag 480  
ggcagcgctg caactccaaa gagccaccta agagggatgt gctcccatga ggcggcaagt 540  
cagcaagggaa aaagggcctt ctctcctgtg cacaggagcc aggattact tatctgttaa 600  
ctt 603

<210> 215

<211> 60

<212> PRT

<213> Homo sapiens

<400> 215

Ser Leu Ala Glu Asp Glu Glu Trp Lys Arg Ile Arg Ser Leu Leu Ser  
1 5 10 15

Pro Thr Phe Thr Ser Gly Lys Leu Lys Glu Met Phe Pro Ile Thr Ala  
20 25 30

Gln Tyr Gly Asp Val Leu Val Arg Asn Leu Arg Arg Glu Ala Glu Lys  
35 40 45

Gly Lys Pro Val Thr Leu Lys Asp Ile Phe Gly Ala  
50 55 60

<210> 216

<211> 603

<212> DNA

<213> Homo sapiens

<400> 216

agcggaaaaac tcaaggaggt atgaaaataa gatgagtctt aattagaaaat gtaaaagaatg 60  
aatctgggga caggtagaaa gtaagatcac agtccgtttc caaggggttag tccactgagt 120  
tcgagcttcc taaaaatggt ctttatctt tatgtacaga aaagacatca caaaattcat 180  
tacaaaatgt cacttactgc tccatgctgg agaaagccat atccttctgg gacttgagtc 240  
tgcacattta actacaggtt ctgatctgtt ttgtgcttag atgttccccca tcattgcccc 300  
gtatggagat gtattggta gaaacttgag gcgaaaagca gagaaaaggca agcctgtcac 360  
cttggaaagag taagtaggag cacagccatg gggttctgag ctgtcatgag cccttccagc 420  
tgcctgccat ggagtcgaca gtcacactgt tggttactc cagtgaccag acaaaaagcag 480  
ggcagcgcgtg caactccaaa gagccaccta agagggagtg gctcccatga ggcggcaagt 540  
cagcaaggga aaagggcctt ctctcctgtg cacaggagcc aggatttact tatctgttaa 600  
ctt 603

<210> 217

<211> 778

<212> DNA

<213> Homo sapiens

<400> 217

cagagcttac atatcttata tcatccacac tcaacacatg ctactgttagt tgtctgataa 60  
tgggtctctg tcttcctatg actgggctcc ttgacctcag aggtgagtct aactcagctt 120  
ggtgtctcca tcaccccccag catagggcca gctccatcac tggcaccaga taaccacctt 180  
ctgagggagt agatggaaga tgattcagca gatagttctg aaagtctgtg gctctttatg 240  
tgtcttactg ggatatgtgg gtttcttgc gcatgtatag tggaaaggacg gtaagagggt 300  
ctgattttaa ttttccatat ctttctccac tcagcatctt tggggcctac agcatggatg 360  
tgattactgg cacatcattt ggagtgaaca tcgactctct caacaatcca caagaccct 420  
ttgtggagag cactaaaaag ttccctaaaaat ttgtttctt agatccatta tttctctcaa 480  
taagtatgtg ggctattatt tctttctctc tttttaaaaa taactgcttt cttgacatat 540  
aattcacata tcgtataatt catccactta aaaggtaaca ttccattgtt ttttaagataa 600  
tcaaaaatat gtatgaccat tactattgtt aactaaaaatg tttttgtcaa tctagagccc 660  
tcacacactt tagctgtcaa caccaccca caaaccacac tgcctctaagc atccaataat 720  
caacttctg cctctataga tttgcctatt ctggacactt catagaaata atatcatt 778

<210> 218

<211> 778

<212> DNA

<213> Homo sapiens

<400> 218

cagagcttac atatcttata tcatccacac tcaacacatg ctactgtagt tgtctgataa 60  
tgggtctctg tcttcctatg actgggctcc ttgacctcag aggtgagtct aactcagctt 120  
ggtgtctcca tcaccccccag catagggcca gctccatcac tggcaccaga taaccacctt 180  
ctgagggagt agatggaaga tgattcagca gatagttctg aaagtctgtg gctctttatg 240  
tgtcttgcact ggatatgtgg gtttcttgct gcatgtatag tggaaaggacg gtaagaggtg 300  
ctgattttaa ttttccatat ctttctccac tcagcatctt tggggcctac agcatggatg 360  
tgattactgg cacatcattt ggagtgaaca tcgactctct caacaatcca caagaccct 420  
ttgtggagag cactaagaag ttccctaaaat ttgggttctt agatccgtt tttctctcaa 480  
taagtatgtg ggcttatttttctc ttttttaaaaa taactgcttt cttgacatata 540  
aattcacata tcgtataatt catccactta aaaggtacaa ttccattgtt ttaagataa 600  
tcaaaaatat gtatgaccat tactattgtt aactaaaatg tttttgtcaa tctagagccc 660  
tcacacactt tagctgtcaa caccccacca caaaccccac tggccctaagc atccaataat 720  
caactttctg cctctataga ttggcctatt ctggacactt catagaaata atatcatt 778

<210> 219

<211> 778

<212> DNA

<213> Homo sapiens

<400> 219

cagagcttac atatcttata tcatccacac tcaacacatg ctactgtagt tgtctgataa 60  
tgggtctctg tcttcctatg actgggctcc ttgacctcag aggtgagtct aactcagctt 120  
ggtgtctcca tcaccccccag catagggcca gctccatcac tggcaccaga taaccacctt 180  
ctgagggagt agatggaaga tgattcagca gatagttctg aaagtctgtg gctctttatg 240  
tgtcttgcact ggatatgtgg gtttcttgct gcatgtatag tggaaaggacg gtaagaggtg 300  
ctgattttaa ttttccatat ctttctccac tcagcatctt tggggcctac agcatggatg 360  
tgattactgg cacatcattt ggagtgaaca tcgactctct caacaatcca caagaccct 420  
ttgtggagag cactaagaag ttccctaaaat ttgggttctt agatccatta tttctctcaa 480  
taagtatgtg ggcttatttttctc ttttttaaaaa taactgcttt cttgacatata 540  
aattcacata tcgtataatt catccactta aaaggtacaa ttccattgtt ttaagataa 600  
tcaaaaatat gtatgaccat tactattgtt aactaaaatg tttttgtcaa tctagagccc 660  
tcacacactt tagctgtcaa caccccacca caaaccccac tggccctaagc atccaataat 720  
caactttctg cctctataga ttggcctatt ctggacactt catagaaata atatcatt 778

<210> 220

<211> 778

<212> DNA

<213> Homo sapiens

<400> 220

cagagcttac atatcttata tcatccacac tcaacacatg ctactgtagt tgtctgataa 60  
tgggtctctg tcttcctatg actgggctcc ttgacctcag aggtgagtct aactcagctt 120  
ggtgtctcca tcaccccccag catagggcca gctccatcac tggcaccaga taaccacctt 180  
ctgagggagt agatggaaga tgattcagca gatagttctg aaagtctgtg gctctttatg 240  
tgtcttgcact ggatatgtgg gtttcttgct gcatgtatag tggaaaggacg gtaagaggtg 300  
ctgattttaa ttttccatat ctttctccac tcagcatctt tggggcctac agcatggatg 360  
tgattactgg cacatcattt ggagtgaaca tcgactctct caacaatcca caagaccct 420  
ttgtggagag cactaagaag ttccctaaaat ttgggttctt agatccatta tttctctcaa 480  
taagtatgtg ggcttatttttctc ttttttaaaaa taactgcttt cttgacatata 540  
aattcacata tcgtataatt catccactta aaaggtacaa ttccattgtt ttaagataa 600  
tcaaaaatat gtatgaccat tactattgtt aactaaaatg tttttgtcg tctagagccc 660  
tcacacactt tagctgtcaa caccccacca caaaccccac tggccctaagc atccaataat 720  
caactttctg cctctataga ttggcctatt ctggacactt catagaaata atatcatt 778

<210> 221

<211> 670

<212> DNA

<213> Homo sapiens

<400> 221

agaagggtgcc attgatctca ctgctgttagt ggtgttcct atgtatagac ctgccttgc 60  
tcagtcggcc gcctgaaaga agggcaaaca tgataaaaagg aatgggttcc agttgagaat 120  
catgatgttc ttattcttat tactggtaga gaaaattata attgctccag gtaaagtgg 180  
catttcaat gatttcctt tggcggttt gttttccca cagactctt tccattcctt 240  
accccagttt ttgaagcatt aaatgtctt ctgtttccaa aagataccat aaattttta 300  
agtaaatctg taaacagaat gaagaaaaagt cgcccaaacg acaaacaaaaa ggtaaaatct 360  
gatgggtggtt aaatgacgt gtttaggttt tgataaaattt agattttata cacatgatag 420  
agcatgtatc tgtatTTTA aaaataaaaga cagagaactt atgttttagaa caagagaagc 480  
catttggtag aaataaaagaa ggagattggg gaaggagatg agaatgagtc agagagatag 540  
catttaaaac ttgaaatcag gcacaacaat tagtatgtca tgatataaaac agtatttgaga 600  
taaaattttt ccacttctct tcccttaat aaattgtcaa aggataaaagt ttctgtttt 660  
aaaatatatt 670

<210> 222

<211> 670

<212> DNA

<213> Homo sapiens

<400> 222

agaagggtgcc attgatctca ctgctgttagt ggtgttcct atgtatagac ctgccttgc 60  
tcagtcggcc gcctgaaaga agggcaaaca tgataaaaagg aatgggttcc agttgagaat 120  
catgatgttc ttattcttat tactggtaga gaaaattata attgctccag gtaaagtgg 180  
catttcaat gatttcctt tggcggttt gttttccca cagactctt tccattcctt 240  
accccagttt ttgaagcatt aaatgtctt ctgtttccaa aagataccat aaattttta 300  
agtaaatctg taaacagaat gaagaaaaagt cgcccaaacg acaaacaaaaa ggtaaaatct 360  
gatgggtggtt aaatgacgt gtttaggttt tgataaaattt agattttata cacatgatag 420  
agcatgtatc tgtatTTTA aaaataaaaga cagagaactt atgttttagaa caagagaagc 480  
catttggtag aaatacagaa ggagattggg gaaggagatg agaatgagtc agagagatag 540  
catttaaaac ttgaaatcag gcacaacaat tagtatgtca tgatataaaac agtatttgaga 600  
taaaattttt ccacttctct tcccttaat aaattgtcaa aggataaaagt ttctgtttt 660  
aaaatatatt 670

<210> 223

<211> 826

<212> DNA

<213> Homo sapiens

<400> 223

agataaaagta cttttaggat cattcaaggc acacacccat aacactgagt atgtaagaca 60  
gaaatgtct ctctggaaat tacagcagtg ctgggtctgg gatgccatga tgaggagtgt 120  
gtggcccaaca atcatgtaga ctttggggaa acctggatta aaatgatttt gcatcatcct 180  
ggccctgtat aagatacata tcagaatgaa aaccactccc agtgtgactt tgaattgctt 240  
ttccatTTTT tcttcttggg attagagagc ttcaatttga tttcatctaa gctgtgatgt 300  
tgtacgttga cctgatttac cttttatgtc tttctctcc tttcagctct gtctgatctg 360  
gagctcgcag cccagtcaat aatcttcatt tttgtggct atgaaaccac cagcagtgtt 420  
ctttccctca ctttatatga actggccact caccctgatg tccagcagaa actgcaaaag 480  
gagattgatg cagtttgcc caataaggtg agggatgac ccctggagat gaagggaga 540  
ggtgaagcct tagaaaaat gcctcctcac cactccccag gagaattttt ataaaaaagca 600  
taatcactga ttcccttact gacataatgt aggaagcctc tgaggagaaa aacaaggaa 660  
gaaacataga gaacgggttgc tactggcaga agcataagat,tttgcataaa tattgtggc 720  
cctggttcac ctgtttactg ttatcacaat aatgtaagt aaaaaaaaaa aaaaaaaaaa 780  
aaaaaaaaa aggagtgtgg cgagaagatg gccaaacagg aacagc 826

<210> 224

<211> 826

<212> DNA

<213> Homo sapiens

&lt;400&gt; 224

agataaaagta cttttaggat cattcaaggc acacacccat aacactgagt atgtaagaca 60  
gaaatgcct ctctggaaat tacagcagtg ctgggtctgg gatgccatga tgaggagtgt 120  
gtggcccaca atcatgtaga ccttggaaa acctggatta aaatgattt gcgtcatcct 180  
ggccctgtat aagatacata tcagaatgaa aaccactccc agtgtgactt tgaattgctt 240  
ttccatttt tcttcttggg attagagagc ttcaacttaga tttcatctaa gctgtgatgt 300  
tgtacgttga cttgatttac ctaaaatgtc tttcctctcc tttcagctct gtctgatctg 360  
gagctcgca cccagtcaat aattttcatt tttgtggct atgaaaccac cagcagtgtt 420  
ctttccttca ctttatataa actggccact caccctgatg tccagcagaa actgaaaaag 480  
gagattgatg cagtttgcc caataaggtg agggatgac ccctggagat gaaggaaaga 540  
ggtgaaggcct tagaaaaat gcctcctcac cactccccag gagaattttt ataaaaaagca 600  
taatcaactga ttcccttact gacataatgt aggaagcctc tgaggagaaa aacaaaggaa 660  
gaaacataga gaacggttgc tactggcaga agcataagat ctttgcataa tattgctggc 720  
cctggttcac ctgtttactg ttatcacaat aatgctaagt aaaaaaaaaa aaaaaaaaaa 780  
aaaaaaaaaa aggagtgtgg cgagaagatg gccaaacagg aacagc . . . . . 826

&lt;210&gt; 225

&lt;211&gt; 616

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 225

caggcctggc acagagtcag tgctccataa atatttgtt aaacgatgga tggtgagtgc 60  
ttttactatac cagattttac ccagcttata gattaagtat gaagagttca agatacatgg 120  
tgttaagagt cgtttttata tgcttgcaaa gcattttgt catatttttt ctactttgct 180  
tccatctttt ctttttccac ttcattttatt aatttccat atgcttgcattt aactattgca 240  
gatccccctt aaatttagaca cgcaaggact tcttcaacca gaaaaaccca ttgttctaaa 300  
ggtggattca agagatggaa ccctaagtgg agaatgagtt attctaaggaa tttctacttt 360  
ggtcttcaag aaagctgtgc cccagaacac cagagatttc aacttagtca ataaaaacctt 420  
gaaataaaaga tggcttaat ctaatgtact gcatgagtag ttggtgattt tgtacattca 480  
ttgagctctc ccagagtcgt ttttagtgcattt gtgttgcattt gtgttgcattt ggaggtgacc 540  
agtagtaatgtca cagataggtt gactcagttt ctctgcttct cataggacta cctctaccca 600  
cctctagttt gcatat . . . . . 616

&lt;210&gt; 226

&lt;211&gt; 624

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 226

gattaagctt ttcatgattt ctcatagaac atgaactcaa aagaggtcag caaagggttg 60  
tgtgcattt tttgttattt gctgcagctt tagccctgcc tccttctcca gcacataaaat 120  
ctttcagcag cttggctgaa gactgtgtt cagggcaggg aagctccagg caaacagccc 180  
agcaaaacagc agcaactcagc taaaaggaag actcacagaa cacagttgaa gaagggaaagt 240  
ggcgatggac ctcatcccaa atttggcggt ggaaacctgg cttctcttgg ctgtcagcct 300  
ggtgctccctc tatctgttag taactgtcca aactcccttc ttgttgcattt tggacttggg 360  
gtgctaatcg ggcccccttc cccttatctt ttttgaagat caaaagagat gttcaaggag 420  
aagtagctga agtggatggac gctacaaacg catagaagttt attattatct tatgcagatc 480  
tatgaatgaa taaaataagca ttctccat ccaccccttca attttggtga cttaggagggt 540  
tttagggacac catttggtag tggaaatgtt ttgattagct tagatctgac gaagactaat 600  
caatgaaaac atggcagcgg caga . . . . . 624

&lt;210&gt; 227

&lt;211&gt; 624

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 227

gattaagctt ttcatgattt ctcatagaac atgaactcaa aagaggtcag caaagggttg 60

tgtgcgattc tttgttattt gctgcagcta tagccctgcc tccttctcca ycaaaaaaaa 120  
 ctttcagcag ctggctgaa gactgctgtg cagggcaggg aagctccagg caaacagccc 180  
 agcaaaacagc agcactcagc taaaaggaag actcacagaa cacagttgaa gaaggaaaagt 240  
 ggcgtatggac ctcatccaa atttggcggt ggaacacctgg cttctcctgg ctgtcagcct 300  
 ggtgtctctc tatctgttag taactgtcca aactcctctc ttgtttccct tggacttggg 360  
 gtgctaatacg ggcccccttt cccttatctg ttttaagat caaaagagat gttcaaggag 420  
 aagtagctga agtgttggac gctacaaacg catagaagtt attattatct tatacagatc 480  
 tatgaatgaa taaataagca ttctcccat ccacccctta attttggtga ctaggagggt 540  
 ttagggacag catttggtag tggaaatgat ttgatttagct tagatctgac gaagactaat 600  
 caatgaaaaac atggcagcgg caga 624

<210> 228  
<211> 626  
<212> DNA  
<213> Homo sapiens

<400> 228  
 gaacagttcc ctaccacgtg gagcatttgc aattaaaagg agactgagat atagaggcag 60  
 gagaccacac cagatggctg ggtctccca ctccacccacat acactcagaa 120  
 gaggcttaggc atctaggatc tccattgagc atcttgata tggcttgcca taatatcata 180  
 tacagtcaat aaatattttgt taaataagga tgcccttca atatatttttg tgcaaccatg 240  
 aagatcacca caactaatgt gagaaaaaat gtttctgttg aactctagtc ttttaggccc 300  
 gtgggattta tgaaaagtgc catctttta gctgaggatg aagaatggaa gagaatacag 360  
 tcattgtgt ctccaacctt caccagcggaa aactcaagg aggtatgaaa ataagatgag 420  
 tcttaatttag aaatgttaaag aatgaatctg gggacaggta gaaagtaaga tcacagtccg 480  
 ttccaaggg gtagtccact gagttcgagc ttccaaaaa tggcttttta tctttatgta 540  
 cagaaaagac atcacaaaat tcattacaaa atgtcactta ctgctccatg ctggagaaag 600  
 ccatatcctt ctggacttg agtctg 626

<210> 229  
<211> 60  
<212> PRT  
<213> Homo sapiens

<400> 229  
 Ser Val Phe Thr Asn Arg Arg Ser Leu Gly Pro Val Gly Phe Met Lys  
 1 5 10 15  
 Ser Ala Ile Ser Leu Ala Glu Asp Glu Glu Trp Lys Arg Ile Gln Ser  
 20 25 30  
 Leu Leu Ser Pro Thr Phe Thr Ser Gly Lys Leu Lys Glu Met Phe Pro  
 35 40 45  
 Ile Ile Ala Gln Tyr Gly Asp Val Leu Val Arg Asn  
 50 55 60

<210> 230  
<211> 626  
<212> DNA  
<213> Homo sapiens

<400> 230  
 gaacagttcc ctaccacgtg gagcatttgc aattaaaagg agactgagat atagaggcag 60  
 gagaccacac cagatggctg ggtctccca ctccacccacat acactcagaa 120  
 gaggcttaggc atctaggatc tccattgagc atcttgata tggcttgcca taatatcata 180  
 tacagtcaat aaatattttgt taaataagga tgcccttca atatatttttg tgcaaccatg 240  
 aagatcacca caactaatgt gagaaaaaat gtttctgttg aactctagtc ttttaggccc 300  
 gtgggattta tgaaaagtgc catctttta gctgaggatg aagaatggaa gagaatacag 360  
 tcattgtgt ctccaacctt caccagcggaa aactcaagg aggcatgaaa ataagatgag 420

tcttaattag aaatgtaaaag aatgaatctg gggacaggta gaaagtaaga tcacagtccg 480  
tttccaagggt gtagtccact gagttcgagc ttccaaaaa tggctttta tctttatgtta 540  
cagaaaagac atcacaaaaat tcattacaaa atgtcactta ctgctccatg ctggagaaag 600  
ccatatcctt ctgggacttg agtctg 626

<210> 231  
<211> 626  
<212> DNA  
<213> Homo sapiens

<400> 231  
gaacagttcc ctaccacgtg gagcatttc aattaaaagg agactgagat atagagggcag 60  
gagaccacac cagatggctg ggtctccccca ctccccacccc cgccccacat acactcagaa 120  
gaggcttaggc atctaggatc tccattgagc atcttgaata tggcttgcca taatatcata 180  
tacagtcaat aaatatttgt taaataagga tgccttca atatattttg tgcaaccatg 240  
aagatccaca caactaatgt gagaaaaaat gtttctgttg aactctagtc tttaggccca 300  
gtgggattna tgaaaaagtgc catctcttta gctgaggatg aagaatggaa gagaatacgg 360  
tcattgtgt ctccaacctt caccagcgga aaactcaagg aggtatgaaa ataagatgag 420  
tcttaattag aaatgtaaaag aatgaatctg gggacaggta gaaagtaaga tcacagtcc 480  
tttccaagggt gtagtccact gagttcgagc ttccaaaaa tggctttta tctttatgtta 540  
cagaaaagac atcacaaaaat tcattacaaa atgtcactta ctgctccatg ctggagaaag 600  
ccatatcctt ctgggacttg agtctg 626

<210> 232  
<211> 623  
<212> DNA  
<213> Homo sapiens

<400> 232  
attggacatg atagcttagat ttgtttcagg aaaacatcct gctttccaag gattnagatg 60  
aatgttttg ttcaactgggt actcaggtaa cacgtcttca agaagccata gggaggttga 120  
gggagggaaag tcaagaagggt aggttggaga ctgcactttt gatttacttc tgacttcacg 180  
agtcactttc tgcgaaagaa atctctccctt ttgcttctag caccgactag atttccttca 240  
gctgtatgatt gactcccaga attcgaaaga aactgagtc cacaaggta accaaggagt 300  
gcttctgagg gctactggcg gggacactaa gagggaggc cttgttctga aaatgtgcag 360  
gaagtattcc aggaagatga gaatttttc cacatagcag aacaacacac atttagatgt 420  
tataaatgggt agctggaggc actttccaga agcccacagg tatagccatg ttccaggctg 480  
aaagggcaac cctaagcaaa cctagaatgc ttggaggaca gtcagtgggtt tgtggatcac 540  
ctacatgaga tcaaatagtcc gttctcagcc tcctccagat ccaccaagtg agaacctcta 600  
cttggaaatt tatatacaaac ata 623

<210> 233  
<211> 616  
<212> DNA  
<213> Homo sapiens

<400> 233  
caggcctggc acagagtca g tgctccataa atattttgtt aaacgatgga tggtgagtgc 60  
ttttactatc cagtatattac ccagcttata gattaagtat gaagagttca agatacatgg 120  
tguttaagagt cgtttttata tgcttgc当地 gcattttgtt catattttt ctactttgct 180  
tccatctttt cttctttcac ttcatattt aattctccat atgcttggttt aactattgtta 240  
gatcccccttg aaatttagaca cgcaaggact tcttcaacca gaaaaacccca ctgttctaaa 300  
ggtggattca agagatggaa ccctaagtg agaatgagttt attctaaggg tttctactttt 360  
ggcttcaag aaagctgtgc cccagaacac cagagattc aacttagtca ataaaacctt 420  
gaaataaaaga tggcttaat ctaatgtact gcatgagtag ttgggtgattt tgtacattca 480  
ttgagcttc ccagagtctg ttttagatgtt tttgcattat gtatgtataaa ggaggtgacc 540  
aggttaagtga cagataggtta gactcagattt ctctgcttcataggacta cctctaccca 600  
cctctagttt gcatat 616

&lt;210&gt; 234

&lt;211&gt; 59

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 234

Met Arg Phe Ala Leu Met Asn Met Lys Leu Ala Leu Ile Arg Val Leu

1 5 10 15

Gln Asn Phe Ser Phe Lys Pro Cys Lys Glu Thr Gln Ile Pro Leu Lys

20 25 30

Leu Asp Thr Gln Gly Leu Leu Gln Pro Glu Lys Pro Thr Val Leu Lys

35 40 45

Val Asp Ser Arg Asp Gly Thr Leu Ser Gly Glu

50 55

&lt;210&gt; 235

&lt;211&gt; 830

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 235

tggtcaccca ccatgtgtac agtaccctgc tagggtccag ggtcatgaaa gtaaataata 60

ccagactgtg cccttgagga actcacctct gctaaggaa acaggcacag aaacccacaa 120

gggtggtaga gagggaaatag gacaatagga ctgtgtgagg gggataggag gcacccagag 180

gaggaaatgg ttacatctgt gtgaggaggt tggtaaggaa agacttaat agaagggtc 240

tgtctggctg ggcttgcaag gatgtgttagg agtcatctag ggggcacaag tacactccag 300

gcagaggaa ttgcatgggt aaagatctgc agttgtggct tggggatg gattcaagt 360

attctgaaat gaagacagcc atggaaacaa gggcagggtga gaggatattt aagaggctc 420

atgcaatgg ctccacttca gtttctgata agaactcagg ttccgtggac tccctgataa 480

aactgattaa gttgtttatg attccccata gaatatgaac tcaaaggagg taagcaaagg 540

gggtgtgtcg attctttgct accggctgca gctgcagccc cacctccttc tccagcacat 600

aaacatttca gcagcttgac ctaagactgc tgtgcagggc agggatgctc cagggcagaca 660

gcccgcaaa caacagcaca cagctgaaag taagactcag aggagacagt tgaagaaggc 720

aagtggcgtt ggacctcatc ccaaattttgg cggtggaaac ctggcttctc ctggctgtca 780

gcctggtgct ccttatctg tcagtaactg tccagattcc ttcctctgt 830

&lt;210&gt; 236

&lt;211&gt; 775

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 236

ggagtacct gatttccag gtgctgtctg tcacccttt ctttactag gaaaggaaac 60

tccttgaccc cttgcgttc tcaagtgagg caatgcctcg ccctgttcg gcttgcac 120

agcacgtgc acccaactgtc ctgcaccac tgcacccac tgcacccac tgcacccac 180

gtacactcaga tgaaaatgc gaaattaccc gtcttcgtg tcaactcacac tggagctgt 240

agaccggagc tggccatctt ggctccaccc cccgagttt ggctttat 300

tgaaaatgtt ttatgtgtgg aaggagataa tgccatgcat ttatgacac atattagagg 360

gtctgagaca atgcacgtca taaaagggtca cctaaggaag aaaaaaaaaaa caaggaaaga 420

caactggaaag aacgtgtac tgggagtc tggggccacca aagtctggag aaaaatggta 480

accacaaggc tccacgccta gtttcaactga agacactcgac actagggtgac ttatggatc 540

cttggtagga caagccttga aaagttttac aatagcaaat gtggacgttg tcagaaccaa 600

atgatgtcactgtgttattt gtgtgtgtgt gtcagtgatgtgtgttataaa atcatgacaa 660

ataaaaggcagg ctgtgaagag gggattccca tgctcgatgtg cctgataaca caactatcac 720

aaacgcttgc cgaaccacaca aaggcaatcc caaccttaca caaaa 775

&lt;210&gt; 237

<211> 25  
<212> DNA  
<213> Homo sapiens

<400> 237  
ttgttggaa atgtttgtc ctatc

25

<210> 238  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 238  
acagggagtt gaccttcata cgtt

24

<210> 239  
<211> 33  
<212> DNA  
<213> Homo sapiens

<400> 239  
tcagggtctc tgaaaatttg acacagagtg cta

33

1

49/49

\\\Ntvossius1\Allgemein\Daten-1\sg\sequencelisting\E3103PCT\E3103PCT.APP

1

47/47

1

4/47

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
11 July 2002 (11.07.2002)

PCT

(10) International Publication Number  
**WO 02/053775 A3**

(51) International Patent Classification<sup>7</sup>: C12Q 1/68, C12N 15/00, 5/10, 9/02, C07K 16/00, C12Q 1/26, G01N 33/573, A61K 39/00, 38/00, 48/00

(71) Applicant (for all designated States except US): EPIDAURUS BIOTECHNOLOGIE AG [DE/DE]; Am Neuland 1, 82347 Bernried (DE).

(21) International Application Number: PCT/EP01/15290

(72) Inventors; and

(22) International Filing Date:  
21 December 2001 (21.12.2001)

(75) Inventors/Applicants (for US only): WOJNOWSKI, Leszek [DE/DE]; Ebenaerstrasse 9, 80637 München (DE). HABERL, Michael [DE/DE]; Waxensteinstrasse 24, 82347 Bernried (DE). HUSTERT, Elisabeth [DE/DE]; Geschwister Scholl Ring 2, 82110 Germering (DE).

(25) Filing Language: English

(74) Agent: VOSSIUS & PARTNER; Siebertstrasse 4, 81675 Munich (DE).

(26) Publication Language: English

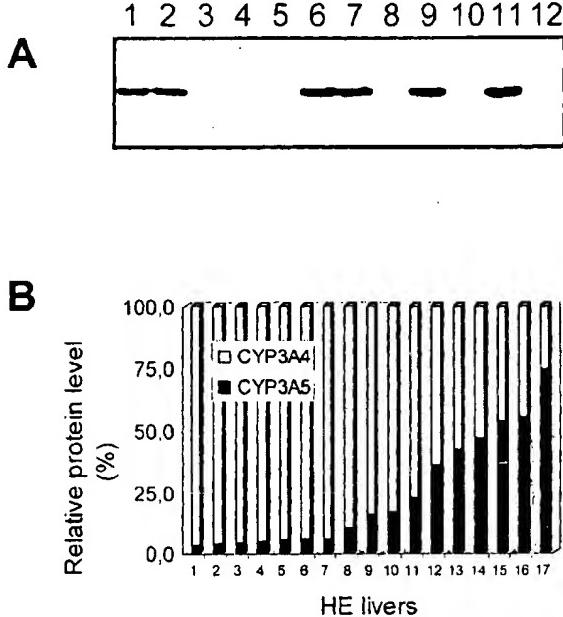
(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

[Continued on next page]

(54) Title: IDENTIFICATION OF GENETIC DETERMINANTS OF POLYMORPHIC CYP3A5 EXPRESSION



**WO 02/053775 A3**



(57) Abstract: The present invention relates to a polymorphic CYP3A5 polynucleotide. Moreover, the invention relates to genes or vectors comprising the polynucleotides of the invention and to a host cell genetically engineered with the polynucleotide or gene of the invention. Further, the invention relates to methods for producing molecular variant polypeptides or fragments thereof, methods for producing cells capable of expressing a molecular variant polypeptide and to a polypeptide or fragment thereof encoded by the polynucleotide or the gene of the invention or which is obtainable by the method or from the cells produced by the method of the invention. Furthermore, the invention relates to an antibody which binds specifically the polypeptide of the invention. Moreover, the invention relates to a transgenic non-human animal. The invention also relates to a solid support comprising one or a plurality of the above mentioned polynucleotides, genes, vectors, polypeptides, antibodies or host cells. Furthermore, methods of identifying a polymorphism, identifying and obtaining a pro-drug or drug or an inhibitor are also encompassed by the present invention. In addition, the invention relates to methods for producing a pharmaceutical composition and methods of diagnosing a disease. Further, the invention relates to a method of

detection of the polynucleotide of the invention. Furthermore, comprised by the present invention are a diagnostic and a pharmaceutical composition. Even more the invention relates to uses of the polynucleotides, genes, vectors, polypeptides or antibodies of the invention. Finally, the invention relates to a diagnostic kit.



(84) **Designated States (regional):** ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

(88) **Date of publication of the international search report:**  
4 September 2003

**Published:**  
— with international search report

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

## INTERNATIONAL SEARCH REPORT

Int	Application No
PCT/EP 01/15290	

A. CLASSIFICATION OF SUBJECT MATTER					
IPC 7	C12Q1/68	C12N15/00	C12N5/10	C12N9/02	C07K16/00
	C12Q1/26	G01N33/573	A61K39/00	A61K38/00	A61K48/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C12Q C07K G01N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

BIOSIS, EPO-Internal, EMBL, MEDLINE, EMBASE, PAJ, WPI Data, CHEM ABS Data

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PAULUSSEN AIMEE ET AL: "Two linked mutations in transcriptional regulatory elements of the CYP3A5 gene constitute the major genetic determinant of polymorphic activity in humans." PHARMACOGENETICS, vol. 10, no. 5, July 2000 (2000-07), pages 415-424, XP008015489 ISSN: 0960-314X cited in the application the whole document abstract page 415, line 15 - line 18	1,3-20, 22-24, 29, 33-35, 38,44
Y	---	1-39,43, 44 -/-

 Further documents are listed in the continuation of box C. Patent family members are listed in annex.

## ° Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

25 March 2003

Date of mailing of the International search report

03 JUL 2003

Name and mailing address of the ISA  
 European Patent Office, P.B. 5818 Patentlaan 2  
 NL - 2280 HV Rijswijk  
 Tel: (+31-70) 340-2040, Tx. 31 651 epo nl  
 Fax: (+31-70) 340-3016

Authorized officer

Tuynman, A

## INTERNATIONAL SEARCH REPORT

Intr	Application No
	PCT/EP 01/15290

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 00 39332 A (JANSSEN PHARMACEUTICA NV ;PAULUSSEN AIMEE DYMPHNE CATHER (BE); ARM) 6 July 2000 (2000-07-06)  the whole document	1-20, 22-24, 29, 33-35, 38,44  1-39,43, 44
Y	WO 98 44939 A (WANG REGINA W ;LU ANTHONY Y H (US); MERCK & CO INC (US)) 15 October 1998 (1998-10-15) example 1	1-39,43, 44
X	WO 98 24914 A (INST NAT SANTE RECH MED ;GENSET SA (FR); TCHOUMAKOV ILIA (FR); BAC) 11 June 1998 (1998-06-11) fragment2.FM0x nucleotides 3450-3460 figure 10	1,3-8, 11-18, 34,44
P,X	HUSTERT ELISABETH ET AL: "The genetic determinants of the CYP3A5 polymorphism." PHARMACOGENETICS, vol. 11, no. 9, December 2001 (2001-12), pages 773-779, XP008015487 ISSN: 0960-314X the whole document	1-35,38, 43,44

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/EP 01/15290

### Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1.  Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
  
2.  Claims Nos.: 41 (fully) ; 1-40,42-44 (in part)  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:  
see FURTHER INFORMATION sheet PCT/ISA/210
  
3.  Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

### Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1.  As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
  
2.  As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
  
3.  As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
  
4.  No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:  
1-39, 43, 44 partially

Remark on Protest

- The additional search fees were accompanied by the applicant's protest.  
 No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 41 (fully) ; 1-40,42-44 (in part)

The search has been not been carried out for claim 41 and those parts of claims 1-40 and 42-44 that refer back to accession numbers, because a molecule which is only identified by reference to a disclosure outside the application filed, which is not a sequence listing according to Rule 5.2, cannot fulfill the requirement of sufficiency of disclosure of Article 5 PCT.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

Inventions 1-102: Claims: 1-39,43,44 (partially)

Each of the polynucleotides comprising a polynucleotide according to SEQ 54,56,58,62,64,66,68,70,72,74, 76,78,80,84,86,90,92,94,96,98,100,102,106,108,110,113,114,115, 116,117,118,119,120,121,122,123,124,125,126,128,129,130,131, 133,134,135,136,137,138,139,140,142,143,149,151,153,155,157,1 59,161,163,165,169,171,173,175,177,179,181,183,185,187,189,19 3,195,197,199,201,207,208,209,210,211,212,213,214,216,218,219 220,221,222,223,224,225,226,227,228,231,232,233,235, or 236, respectively, and methods and uses involving these.

Inventions 103-105: claims 1-40,42-44 (partially)

Each of the polynucleotides comprising a polynucleotide according to SEQ 82,88 or 112, respectively, and methods and uses involving these.

Inventions 106-108: claims 1-40,42-44 (partially)

A polynucleotide encoding a polypeptide according to SEQ ID 127,132, or 141, respectively, and methods and uses involving these.

Inventions 109-111: claims 1-39,43,44 (partially)

A polynucleotide encoding a polypeptide according to SEQ ID 215,229, or 234, respectively, and methods and uses involving these.

Invention 112: 38-40,42-44 (partially)

A polynucleotide comprising a polynucleotide according to SEQ ID NO: 104 and the use thereof.

Invention 113: Claims 38-40,42-44 (partially)

A polynucleotide encoding a polypeptide comprising SEQ ID 145 and the use thereof.

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

Inte	nal Application No
PCT/EP 01/15290	

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 0039332	A 06-07-2000	AU 1878300 A		31-07-2000
		CA 2355658 A1		06-07-2000
		CN 1331756 T		16-01-2002
		EP 1141398 A1		10-10-2001
		WO 0039332 A1		06-07-2000
		JP 2002533136 T		08-10-2002
<hr/>				
WO 9844939	A 15-10-1998	EP 1011708 A1		28-06-2000
		JP 2000513742 T		17-10-2000
		US 6300476 B1		09-10-2001
		WO 9844939 A1		15-10-1998
<hr/>				
WO 9824914	A 11-06-1998	FR 2756845 A1		12-06-1998
		AU 746293 B2		18-04-2002
		AU 5327198 A		29-06-1998
		EP 0941341 A1		15-09-1999
		WO 9824914 A1		11-06-1998
		JP 2001505430 T		24-04-2001
		US 6551792 B1		22-04-2003
<hr/>				